П. И. ПЕНИН

СИСТЕМЫ ПЕРЕДАЧИ ЦИФРОВОЙ ИНФОРМАЦИИ



П. И. ПЕНИН

СИСТЕМЫ ПЕРЕДАЧИ ЦИФРОВОЙ ИНФОРМАЦИИ

УЧЕБНОЕ ПОСОБИЕ ДЛЯ ВУЗОВ

Допущено Министерством высшего и среднего специального образования СССР в качестве учебного пособия для студентов радиотехнических специальностей вузов



МОСКВА «СОВЕТСКОЕ РАДИО» 1976

6Ф7
П 25

УДК 621.391.1

Пенин П. И. Системы передачи цифровой информации. Учебное пособие для вузов. М., «Сов радио», 1976, 368 с.

В книге излагаются принципы, методы и способы создания систем передачи цифровой информации Большое внимание уделяется постановке рассматриваемых задач и получению соотношений для инженерных расчетов и оценок.

Рассматриваются особенности систем, критерии оценки качества их работы и предельные возможности при передаче цифровой информации, а также вопросы представления информации в цифровой форме и особенности передачи ее по каналам как с постоянными, так и со случайными параметрами. Анализируются различные методы повышения качества передачи цифровой информации по каналам с помехами. Излагается методика инженерного расчета основных параметров цифровых линий связи и приводятся примеры.

Книга является учебным пособием для студентов радиотехнических специальностей вузов, а также может быть полезна специалистам в области теории и техники передачи цифровой информации.

70 рис., 18 табл., библ. 161 назв.

Рецензенты:

Кафедра радиосистем Московского электротехнического института связи и канд. техн. наук В. Ф. Уханов

Редакция кибернетической литературы

П $\frac{30501\text{-}029}{046(01)\text{-}76}$ 72-76

ПРЕДИСЛОВИЕ

Книга представляет собой учебное пособие по курсу «Системы передачи цифровой информации» и предназначена в первую очередь для студентов радиотехнических специальностей вузов. Основой книги явился курс лекций, который читался автором в Московском энергетическом институте в 1969—1973 гг. Главная цель, которую поставил автор, состояла в том, чтобы изложить в систематизированном и доступном виде ряд важных вопросов передачи цифровой информации.

В книге излагаются идеи, принципы и методы, которые лежат в основе построения современных цифровых информационных систем. При рассмотрении вопросов особое внимание обращено на постановку задач, а также на обсуждение исходных данных, принимаемых идеализаций и допущений. По возможности автор стремился не перегружать материал пособия математическими выкладками и доказательствами. Поэтому в ряде случаев строгому обоснованию рассматриваемых результатов уделяется меньшее внимание, чем их прикладной стороне. Всюду, где появлялась возможность, изложение доводилось до соотношений, позволяющих проводить инженерные расчеты и оценки.

Автор не ставил перед собой задачу рассмотреть все наиболее существенные вопросы цифровой передачи информации. Такие вопросы, как синхронизация в цифровых системах связи, учет межсимвольных искажений, кодирование и декодирование информации и ряд других, рассмотрены очень кратко, а некоторые вопросы вообще не затронуты.

Несмотря на это автор надеется, что пособие облегчит студентам знакомство с многочисленной и непрерывно увеличивающейся литературой по цифровым системам связи, позволит им быстрее и правильнее ориентироваться в этой литературе при решении тех или иных вопросов в процессе учебно-исследовательской работы, курсового и дипломного проектирования.

Книга может быть полезной и инженерам-практикам, желающим хотя бы частично восполнить пробел в области основ статистической теории связи.

Для закрепления материала и контроля за его усвоением в каждой главе имеются вопросы и задачи. Часть задач сформулирована так, чтобы читатель мог не только проверить усвоение материала, но и получить некоторые дополнительные сведения.

Автор весьма признателен проф. А. И. Дымовой и сотрудникам возглавляемой ею кафедры за внимание к работе, а также кандидатам технических наук В. С. Кузькину и В. Ф. Уханову за деловую критику и замечания, которые были учтены при окончательной подготовке рукописи к печати.

Глава 1

Общие сведения о системах передачи цифровой информации

1.1. ВВЕДЕНИЕ

В различных областях деятельности современного общества все больше и больше приходится заниматься вопросами сбора, преобразования, хранения, передачи, обработки и использования разнообразной научной, технической, экономической и т. п. информации. Установлено, что объем информации, необходимой для деятельности современного общества, возрастает приблизительно пропорционально квадрату развития производительных сил.

Непрерывно возрастающие потоки информации становится немыслимо обрабатывать без применения средств электронно-вычислительной техники.

Развитие автоматизированных средств обработки информации с применением электронных вычислительных машин (ЭВМ) выдвигает, в свою очередь, ряд новых задач, часть которых начинает выходить за рамки чисто научных и технических проблем. Степень экономического развития любого крупного современного государства уже недостаточно характеризовать только такими традиционными показателями, как выплавка стали, добыча нефти и угля, выработка электроэнергии и т. п. К ним добавляются такие важные показатели, как: уровень развития и использования информационно-измерительной и электронно-вычислительной техники, насыщенность различных отраслей автоматизированными системами управления, системами передачи информации и т. д.

В Советском Союзе уделяется очень большое внимание созданию различных автоматизированных систем управления с применением средств электронно-вычислительной техники. Такие системы находят широкое применение в промышленности, в сфере материально-технического снабжения предприятий, на транспорте, в торговле и других областях народного хозяйства. Важную роль играют автоматизированные системы управления в ракетной и космической технике, в радиолокации и т. п. Замечательными примерами здесь являются решение задачи по доставке лунного грунта на Землю с помощью автоматических систем «Луна 16» и «Луна 20», исследование Луны с помощью систем «Луноход 1» и «Луноход 2», исследование планет Венера и Марс.

Автоматизированные системы управления (АСУ) представляют собой сложные комплексы, в состав которых могут входить различные подсистемы, обеспечивающие выполнение операций по сбору, преобразованию, хранению, передаче, приему, обработке и использованию необходимой информации. Перечисленные операции должны быть взаимосвязаны и согласованы. Действительно, если обработка займет много времени, то полученная информация может устареть и потерять значительную часть своей ценности. При медленной обработке информации управление будет сильно запаздывать, что в ряде случаев совершенно недопустимо, так как может привести к результатам, противоположным ожидаемым. В современных АСУ все операции должны выполняться в реальном масштабе времени, т. е. в соответствии со скоростью изменений состояния управляемых объектов или протекания процессов, связанных с этими объектами.

Обязательной и важной частью любой АСУ являются каналы передачи информации. Качество работы этих каналов в существенной мере определяет качество работы всей АСУ. Для обеспечения высокого качества работы указанных каналов необходимо представлять передаваемую информацию в дискретном виде и использовать соответствующие методы ее передачи.

Системы, с помощью которых обеспечивается передача дискретной информации, часто называют цифровыми, так как передаваемая этими системами информация может рассматриваться как последовательность

чисел, выраженных в удобной для практического применения форме. Цифровые системы передачи информации часто также называют кодовыми системами или системами с кодово-импульсной модуляцией.

Возрастающие потребности в обработке быстро увеличивающихся потоков информации стимулируют развитие систем передачи цифровой информации не только в области управления, но и в других важных областях, таких, например, как телеметрия, многоканальная передача речи, информационные сети с коммутацией сообщений и т. д.

Передача информации в цифровой (дискретной) форме имеет ряд особенностей по сравнению с передачей в аналоговой (непрерывной) форме. Рассмотрим кратко эти особенности.

1. Автоматизированная обработка информации предъявляет повышенные требования к качеству работы канала связи, которое определяется скоростью передачи информации и ее достоверностью.

Принципиальный недостаток аналоговых систем передачи информации заключается в существенном влиянии неидеальности и нестабильности аппаратурных характеристик системы на качество передаваемой информации. Искажения информации, обусловленные этими факторами, часто являются основной причиной существенного ухудшения качества работы системы в целом.

В системах цифровой передачи информации большинство операций сводится к процедурам, в основе которых лежат простейшие логические операции типа «да — нет», «и», «или» и т. п. Поэтому такие важнейшие этапы преобразований сообщений и сигналов, как дискретизация, кодирование, модуляция, усиление, демодуляция, фильтрация, могут быть осуществлены с очень малыми погрешностями.

Большим преимуществом цифровых методов передачи информации по сравнению с аналоговыми является возможность регенерации (восстановления) цифровой последовательности, существенно искаженной различными помехами при передаче по линии связи. Это позволяет значительно ослабить эффект накопления искажений при передаче информации по протяженным линиям связи. Возможность существенного уменьшения влияния аппаратурных погрешностей при цифровой передаче

информации открывает пути к значительному улучшению качественных показателей системы.

2. Важным достоинством передачи информации в цифровой форме является ее универсальность для любых видов передаваемых сообщений (речь, телевизионные изображения, дискретные данные, команды и т. п.). Универсальная форма представления информации позволяет использовать одни и те же каналы связи для передачи сообщений самой различной физической природы.

3. Ключевые режимы работы значительной части электронных устройств, входящих в состав систем передачи цифровой информации, открывают ряд совершенно новых возможностей по фильтрации, преобразованию и усилению сообщений и сигналов. В настоящее время уже известны дискретные устройства, которые выполняют указанные операции. Класс этих устройств, называемых цифровыми фильтрами, непрерывно расширяется. Обладая совершенно уникальными свойствами (высокой стабильностью характеристик, возможностью изменения различных параметров фильтра в процессе работы, высоким быстродействием и т. п.), цифровые фильтры открывают широкие возможности построения высококачественных цифровых систем передачи информации. Выполнение цифровых фильтров на интегральных элементах или в виде интегральных подсистем позволяет унифицировать отдельные элементы и блоки цифровых систем. Унификация и стандартизация отдельных элементов, блоков и подсистем, входящих в состав цифровых систем, позволяет широко применять автоматизацию в производстве аппаратуры и существенно упрощает ее регулировку и настройку.

Цифровые системы принципиально более гибки, чем аналоговые, позволяют применять более совершенные методы передачи и приема информации и способы их реализации. По сравнению с аналоговыми цифровые системы обладают более высокой помехоустойчивостью и надежностью связи, возможностью скрытности работы и т. д.

Большим достоинством цифровой передачи информации является возможность унификации характеристик каналов передачи различного назначения, что позволяет создавать сложные информационные сети, входящие в состав единой автоматизированной системы связи

(ЕАСС). По сравнению с аналоговыми системами передачи информации цифровые системы допускают более простое сочленение их с цифровыми электронными вычислительными машинами.

Итак, наиболее существенными достоинствами цифровых систем передачи информации являются:

1) возможность автоматизации обработки информации;

2) универсальная форма представления сообщений различной физической природы и, как результат этого, гибкость систем, позволяющая, заменив программу работы, применять одно и то же оборудование для разных целей;

3) высокие качественные показатели работы систем;

4) возможность объединения отдельных систем в более крупные системы и комплексы.

Достоинства систем передачи цифровой информации связаны со значительным усложнением схем и технологии изготовления аппаратуры, а также с необходимостью использования более широкой полосы частот по сравнению с аналоговыми системами. Однако в настоящее время усложнение аппаратуры не является принципиальным препятствием, поскольку быстрое развитие современной технологии производства электронной аппаратуры создает широкие возможности реализации цифровых систем на основе новейших достижений в области микроэлектроники, интегральных схем, физики твердого тела и т. п.

Уменьшение размеров, веса, потребляемой энергии и, наконец, стоимости элементов и узлов аппаратуры при непрерывно увеличивающейся их надежности стимулирует применение все более сложных и совершенных схем, обладающих высокими информационными, технологическими и эксплуатационными качественными показателями.

Развитие теории и техники «сжатия информации» позволяет в ряде случаев существенно повышать эффективность использования отводимой полосы частот для цифровых систем.

Цифровые системы передачи информации в настоящее время быстро развиваются и находят все более широкое применение в различных областях.

1.2. ОСНОВНЫЕ ПОКАЗАТЕЛИ КАЧЕСТВА ПЕРЕДАЧИ ЦИФРОВОЙ ИНФОРМАЦИИ

Независимо от того, для каких целей и с помощью каких средств обеспечивается передача цифровой информации, качество передачи этой информации можно характеризовать рядом показателей. К наиболее важным из них относятся достоверность и скорость передачи информации. Рассмотрим суть этих показателей.

Достоверность передачи информации. Передача информации по линиям связи всегда сопровождается неизбежным действием помех и искажений. Это приводит к тому, что принятое сообщение может в какой-то мере отличаться от переданного. Естественно, что при проектировании и создании линий связи всегда стремятся обеспечить такие условия работы, чтобы указанное различие было невелико и не превышало некоторой допустимой величины.

Достоверностью передачи информации принято называть степень соответствия принятого сообщения переданному. Для количественной оценки этого соответствия в случае передачи цифровой информации целесообразно пользоваться отношением числа ошибочно принятых элементов сообщения $M_{\text{ош}}$ к общему числу переданных элементов $M_{\text{общ}}$:

$$k_{\text{ош}} = M_{\text{ош}}/M_{\text{общ}}. \qquad (1.1)$$

Это отношение называется частостью ошибок (или коэффициентом ошибок). При ограниченном времени передачи величина $k_{\text{ош}}$ является случайной и зависит от этого времени. Однако если общее время передачи информации (сеанс связи) значительно превышает длительность отдельного элемента сообщения, а статистические характеристики процесса передачи при этом остаются неизменными (или меняются незначительно), то величина $k_{\text{ош}}$ остается весьма устойчивой и почти не меняется от сеанса к сеансу. Практически такие условия во многих случаях выполняются. Поэтому в первом приближении коэффициент ошибок можно считать близким к вероятности ошибки приема одного элемента сообщения $p_{\text{ош}}$.

Итак, в дальнейшем достоверность передачи цифровой информации будем оценивать величиной вероятно-

сти ошибочного приема одного элемента сообщения. Несмотря на то, что указанный показатель является идеализированным и условия его применимости не всегда строго выполняются, он нашел очень широкое применение при оценке качества передачи цифровой информации из-за своей простоты и удобства.

Скорость передачи информации. При исследовании вопросов передачи цифровой информации приходится сталкиваться с двумя различными понятиями скорости передачи: технической и информационной.

Техническая скорость характеризует быстродействие аппаратуры, входящей в состав передающей части системы связи. Она определяется количеством элементов дискретного сообщения, переданных в секунду. Эта характеристика была предложена в телеграфии около 100 лет назад французским инженером Ж. Бодо. В его честь единица технической скорости была названа бодом.

Техническая скорость передачи определяется величиной

$$R = 1/\tau_0 \ [\text{бод}], \qquad (1.2)$$

где τ_0 — длительность посылки, соответствующей передаче одного элемента дискретного сообщения.

В литературе техническую скорость передачи иногда называют скоростью манипуляции.

В связи с появлением и быстрым развитием теории информации и ее многочисленных приложений возникла необходимость в широком применении понятия информационной скорости передачи. Под этой скоростью понимают количество информации, поступившее по линии связи от источника информации к получателю за одну секунду. Информационная скорость измеряется числом двоичных единиц (бит*) в секунду. Она зависит от ряда факторов: технической скорости передачи, статистических свойств источника, типа канала связи, применяемых сигналов и помех, действующих в этом канале.

Техническую скорость передачи нельзя путать с информационной скоростью, а термин «бод» использовать как синоним термина «бит/с». Количественно эти скорости совпадают только для бинарных симметричных

---

* Термин «бит» происходит от английского названия двоичной единицы binary digit.

линий связи с высокой достоверностью передачи. Только в этом единственном случае можно принять, что 1 бод равен 1 бит/с*. В общем случае информационная скорость не совпадает с технической и может быть как больше, так и меньше ее.

Конкретные значения вероятности ошибки и скорости передачи цифровой информации существенно зависят от типа канала связи, вида сигнала и его энергии, уровня помех в канале, требований к качеству передачи и т. д. Эти вопросы рассмотрены ниже в гл. 3—7.

1.3. ОБЛАСТИ ПРИМЕНЕНИЯ СИСТЕМ ПЕРЕДАЧИ ЦИФРОВОЙ ИНФОРМАЦИИ

В настоящее время передача информации в цифровой форме находит широкое применение в решении самых различных задач связи, телеметрии и управления. Особенно важное значение системы передачи цифровой информации приобрели в таких областях, как передача данных, космическая связь и управление.

1.3.1. Передача данных

Передача данных сравнительно новый вид связи. Его появление относится к началу 50-х годов и обусловлено широким развитием электронных вычислительных машин и их применением в различных системах управления и обработки информации, в частности, в системах управления ракетами, космическими аппаратами, искусственными спутниками Земли, в системах централизации и обработки результатов радиолокационного слежения, в системах управления производством и т. п.

Дать точное определение термина «данные» затруднительно, так как в это понятие в настоящее время вкладывается очень широкий и неоднозначный смысл. Под данными понимается самая различная информация. Как правило, данные имеют дискретную форму. В аналоговой форме передача данных почти не осуществляется и чаще всего такие данные для передачи преобразуются в дискретную форму. Данные можно рассматривать как определенный вид информации, передаваемой на расстоянии средствами связи с целью обработки ее вычислительными машинами или уже обработанной ими. Поэтому данные часто называют «языком машин». Этим подчеркивается та особенность данных, что они не предназначены непосредственно для человека как получателя информации. Осмысливание данных человеком может происходить только после их соответствующей обработки и представления в форме, удобной для окончательного использования. Передача данных, как правило, ве-

* К сожалению, в некоторых публикациях иногда указанное обстоятельство игнорируется, что приводит к неверным выводам.

дется в цифровой форме и основана на тех же принципах, что и передача телеграфных сообщений. Поэтому для ее осуществления широко используют имеющиеся линии телефонной и телеграфной связи.

По сравнению с передачей обычных телеграфных сообщений передача данных требует более высоких скоростей, лучшей достоверности и надежности. Это обстоятельство приводит к ряду трудностей, которые необходимо преодолевать теми или иными техническими способами, позволяющими позысить качество передачи по используемым каналам.

Для телефонных каналов и сетей связи характерны импульсные помехи и кратковременные прерывания связи. Такие помехи (не очень опасные для телефонных переговоров) оказывают существенное влияние на передачу данных. Ошибки при передаче данных могут также возникать из-за неравномерности фазовых характеристик используемых каналов, приводящей к неравномерности характеристик группового времени запаздывания. В результате этого появляются дополнительные переходные процессы и увеличивается взаимное влияние элементов передаваемой последовательности друг на друга (межсимвольная интерференция).

К настоящему времени в ряде стран проведены многочисленные исследования различных мешающих факторов при передаче данных по соответствующим телеграфным и телефонным каналам и накоплен обширный статистический материал, на основе которого построены различные теоретические модели распределения возникающих ошибок. Это позволило разработать ряд методов и способов защиты, позволяющих существенно уменьшать влияние указанных факторов. Разработаны также эффективные способы коррекции частотных характеристик используемых каналов, позволившие значительно повысить достоверность и скорость передачи данных.

Для передачи данных используются различные каналы связи: проводные, кабельные, радиорелейные и прочие радиоканалы, включая каналы спутниковой связи. Каналы связи, используемые для передачи данных, принято делить на коммутируемые и некоммутируемые («арендованные»).

Коммутируемым называют такой канал, который создается из отдельных участков только на время передачи (обычно не более 5—10 мин). По окончании сеанса связи канал ликвидируется. При необходимости вновь установить связь между прежними абонентами канал может быть образован уже из других участков, которые к этому времени оказались свободными. В результате число участков и их длина каждый раз могут быть различными и, следовательно, характеристики коммутируемого канала могут случайно меняться.

Некоммутируемым называют такой канал, который предоставляется абонентам на длительное время или постоянно. Характеристики этого канала постоянны, а уровень помех в нем значительно ниже, чем в коммутируемом канале.

В зависимости от вида используемого канала скорость передачи данных может быть различной. По этому показателю каналы делят на три группы:

1) низкоскоростные (телеграфные) каналы со скоростями передачи от 50 до 200 бит/с;

2) среднескоростные (телефонные) каналы со скоростями передачи от 200 до 4800 бит/с;

3) высокоскоростные (широкополосные) каналы со скоростями в десятки и сотни тысяч двоичных единиц в 1 с (применяются скорости 48 и 230 кбит/с)

Допустимая вероятность ошибки при передаче одного элемента последовательности по современным стандартным телеграфным и телефонным каналам обычно лежит в пределах от 10^{-4} до 10^{-5} (при отсутствии специальных средств защиты от ошибок).

Во всех экономически развитых странах передача данных развивается наиболее быстрыми темпами по сравнению с другими видами связи, а область применения этого вида связи непрерывно расширяется, охватывая почти все сферы деятельности человеческого общества. Это обстоятельство заставляет уделять очень большое внимание вопросам стандартизации параметров систем передачи данных и унификации аппаратуры этих систем в международном масштабе. Разработкой рекомендаций и предложений в области передачи данных занимается ряд международных организаций: Международный консультативный комитет по телефонии и телеграфии (МККТТ), Международная электротехническая комиссия (МЭК), Международная организация по стандартизации (МОС). Разработанные и принятые рекомендации систематически публикуются в специально издаваемых сборниках и, как правило, кладутся в основу создаваемой аппаратуры передачи данных.

Современное состояние техники передачи данных и тенденций ее развития детально рассмотрено в работе [1]. Вопросы теории передачи данных подробно излагаются в работах [2, 3, 4].

1.3.2. Космическая связь

Бурное развитие ракетно-космической техники и непрерывно расширяющиеся программы космических исследований поставили перед техникой радиосвязи ряд совершенно новых задач. К их числу относятся:

обеспечение устойчивой и надежной связи на дальностях в сотни тысяч и сотни миллионов километров;

передача с борта космических объектов на пункты приема большого количества разнообразной информации о состоянии и работе систем и агрегатов этих объектов, а также о различных процессах, происходящих в космическом пространстве и на поверхности исследуемых планет;

управление режимом работы космических объектов путем передачи командной информации с Земли на борт объекта.

В настоящее время успешно используют следующие основные виды систем космической связи:

1 Системы связи «Земля — борт» и «борт — Земля» с искусственными спутниками Земли и орбитальными космическими кораблями и лабораториями.

2. Системы дальней космической связи с автоматическими станциями, запускаемыми в сторону планет солнечной системы.

3. Системы связи с автоматическими станциями и лабораториями, запускаемыми в сторону Луны и на Луну

Как правило, большинство космических систем связи являются совмещенными, т е. представляют собой сложные комплексы, позволяющие одновременно решать задачи связи, телеметрии, измерения параметров движения (траекторные измерения) и управления. Создание подобных систем требует решения ряда сложнейших теорети-

ческих и технических задач, так как системы космической связи должны иметь очень высокие качественные показатели. Эти системы должны работать надежно в течение длительных сроков, измеряемых месяцами и годами; обеспечивать высокую точность передачи информации; обладать высокой энергетической эффективностью; иметь малые веса и габариты и т. д Реализовать такие требования можно только методами и средствами цифровой техники.

Как правило, системы космической связи работают на УКВ и используют несущие частоты сантиметрового и дециметрового диапазонов волн. Поэтому для радиолиний космической связи наиболее характерной помехой является случайная помеха типа «белого шума». Спектральную плотность мощности такой помехи (энергетический спектр) в полосе пропускания приемного устройства можно считать постоянной и равной

$$N_0 = kT_э,$$

где $k = 1{,}38 \cdot 10^{-23}$ Дж/К — постоянная Больцмана; $T_э$ — эквивалентная шумовая температура по шкале Кельвина, обусловленная всеми источниками помех.

Шумовая помеха в системах космической связи создается источниками внутренних (антенно-фидерные устройства и первые каскады приемника) и внешних шумов (радиошумы излучения Солнца, галактические шумы, тепловое излучение Земли, шумы возникающие в атмосфере и т. д.).

Методика определения эквивалентной шумовой температуры $T_э$ изложена в работах [5, 6], и здесь на этих вопросах останавливаться не будем. Отметим только, что применение малошумящих входных устройств в приемниках и тщательный выбор параметров космических систем связи позволяют получить эквивалентную температуру, не превышающую нескольких десятков градусов по шкале Кельвина (20—50 К).

Подробные сведения об особенностях космических систем связи, а также о теории и технике таких систем можно найти в работах [7—9].

1.3.3. Цифровая телеметрия

Телеметрией называют область техники, которая занимается вопросами измерении различных физических величин, характеризующих состояние исследуемых объектов или процессов; передачей результатов этих измерении на расстояние; регистрацией и обработкой полученных данных в пункте приема.

Совокупность технических средств, обеспечивающих измерение, передачу, регистрацию и обработку телеметрической информации, называется телеметрической системой. Иногда телеметрические системы бывают составной частью более крупных систем и комплексов (например, в космических комплексах). Однако имеются области науки, техники и народного хозяйства, в которых телеметрические системы применяются как отдельные самостоятельные системы. К таким областям относятся авиационная и ракетно-космическая техника, океанология и морская техника, промышленность, метеорология, геофизика, медицина и т. п

В зависимости от вида используемого канала связи различают радио, проводные и гидроакустические телеметрические системы.

Характерной особенностью современных телеметрических систем является большое число разнообразных измеряемых физических величин (десятки, сотни и более) и высокая точность измерений. (В ряде случаев погрешность измерения не должна превышать 0,5% полной шкалы изменения измеряемой физической величины.) Информационные потоки в некоторых телеметрических системах могут достигать 10^5—10^6 бит/с.

Узким местом многих применявшихся ранее телеметрических систем было несоответствие между высокой скоростью поступления результатов измерений и низкой скоростью их обработки. Устранение этого несоответствия автоматизацией обработки принятой информации привело к естественной необходимости создания цифровых телеметрических систем. В настояще время цифровые телеметрические системы постепенно вытесняют применявшиеся ранее аналоговые системы [10—12].

1.3.4. Системы дальней цифровой связи и управления

Потребности в надежной и высококачественной передаче значительных потоков информации на дальние расстояния в сотни и тысячи километров непрерывно возрастают.

Ограниченные возможности радиосвязи на длинных, средних и коротких волнах, обусловленные их сравнительно узкими частотными диапазонами, привели к необходимости применять для указанных целей УКВ. УКВ диапазон обладает большой частотной вместимостью и позволяет создавать устойчивые, надежные и высококачественные системы связи. Для увеличения дальности действия УКВ систем связи широко применяются радиорелейные линии с наземными промежуточными ретрансляционными пунктами, а также спутниковые линии связи, в которых ретранслятор размещен на искусственном спутнике Земли. Наряду с такими линиями в последние годы получили заметное развитие тропосферные, ионосферные и метеорные линии связи [13, 14]. Особенно важное значение эти линии связи имеют для решения ряда специальных задач военного характера [15].

Различные каналы передачи цифровой информации по условиям их работы можно разделить на две группы.

К первой группе относятся каналы, в которых за время сеанса связи параметры сигналов при прохождении от передатчика к приемнику не получают каких либо заметных дополнительных изменений. Иными словами, к первой группе относятся каналы с идеальной (или почти идеальной) средой распространения сигналов. В таких каналах основное влияние на качество передачи информации оказывают флюктуационные шумы, а также нестабильность параметров и неидеальность характеристик аппаратуры систем. Подобные каналы связи называются каналами с постоянными параметрами или гауссовскими каналами К ним относятся обычные радиорелейные линии с промежуточными ретрансляторами; УКВ линии связи с дальностью действия в пределах прямой видимости, в частности, спутниковые линии связи с активным ретранслятором; космические линии связи.

Ко второй группе относятся каналы, в которых за время сеанса связи условия распространения сигналов в среде существенно меня-

ются. Эти изменения могут быть вызваны как случайным изменением характеристик среды под действием различных физических факторов, так и случайным перемещением положения диаграммы направленности антенны передатчика или приемника при движении объектов, между которыми устанавливается связь. В некоторых случаях указанные причины могут существовать совместно (например, связь с объектом, движущимся в среде со случайными параметрами).

Каналы второй группы называются каналами со случайными параметрами. Случайные изменения условий распространения сигналов приводят к заметным флюктуациям уровня принимаемого сигнала и времени его запаздывания, что существенно сказывается на качестве принимаемой информации. Поэтому обеспечить высокую достоверность передачи информации в каналах со случайными параметрами значительно труднее, чем в каналах с постоянными параметрами.

Важной особенностью требований к современным и перспективным системам передачи цифровой информации (особенно к системам передачи данных) является требование стандартизации показателей качества передачи. Так как показатели различных систем могут существенно различаться, то за основу при создании любой системы принимается стандартный телефонный канал передачи данных с показателями, устанавливаемыми для такого канала в соответствии с рекомендациями МККТТ. Это позволяет обеспечить сопряжение различных линий и каналов связи в единую систему (когда в этом возникает необходимость), называемую сетью связи.

Без стандартизации наиболее важных показателей качества передачи информации по различным линиям связи немыслимо создание единой общегосударственной автоматизированной системы связи, сбора и обработки разнообразной информации (ЕАСС). Как известно, в нашей стране ведутся обширные исследования и разработки, связанные с созданием этой чрезвычайно важной и необходимой системы.

1.4. СТРУКТУРНАЯ СХЕМА ЦИФРОВОЙ ЛИНИИ СВЯЗИ

Специфика различных областей применения систем передачи цифровой информации требует существенно различного подхода к реализации таких систем. Система передачи данных по телефонным каналам совершенно не похожа на систему космической связи или на цифровую систему тропосферной связи ни по техническому исполнению, ни по параметрам. Однако в принципах построения и назначения отдельных устройств самых разных систем много общего. Это дает возможность рассматривать цифровые системы связи с общих позиций, отвлекаясь от специфики их применения и конкретных условий работы. Такой подход позволяет

выяснить ряд важных положений, характерных для любой системы передачи цифровой информации.

В дальнейшем под системой передачи цифровой информации будем понимать упорядоченную совокупность технических средств, с помощью которых осуществляется передача информации в цифровой форме из одного пункта пространства в другой. Такая система вместе с физической средой, в которой распространяются содержащие передаваемую информацию сигналы, образует цифровую линию связи, структурная схема которой приведена на рис. 1.1.

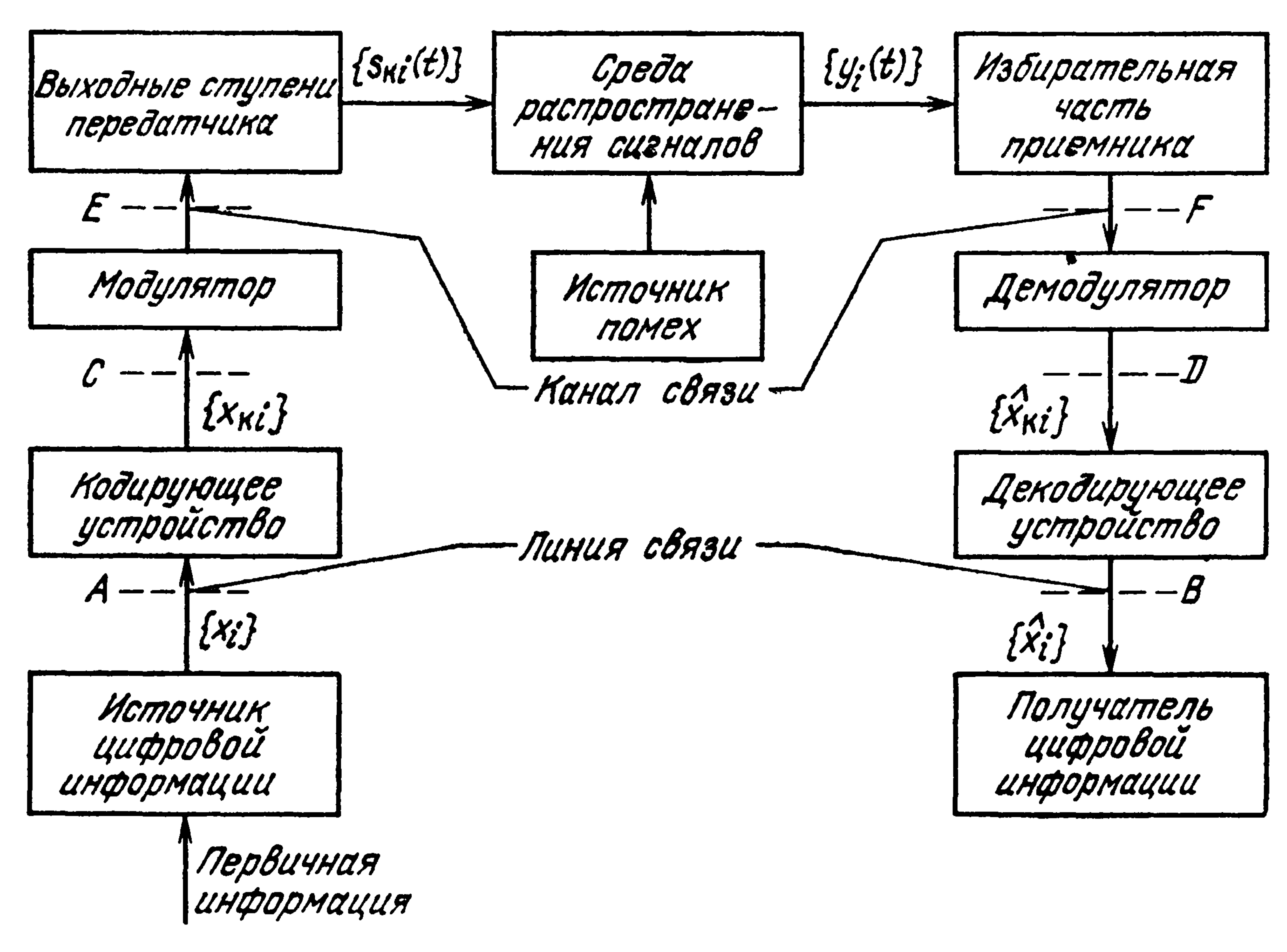


Рис. 1.1.

По цифровым линиям связи можно передавать сообщения самой различной физической природы: цифровые данные, полученные от ЭВМ, речь, тексты телеграмм, команды управления, результаты измерений различных физических величин с телеметрируемых объектов и т. д. Естественно, что все эти сообщения предварительно должны быть преобразованы в электрические колебания, сохраняющие все свойства исходных сообщений, а затем унифицированы, т. е. представлены в цифровой форме, удобной для последующей передачи *.

* Вопросы представления различных сообщений в цифровой форме подробно рассматриваются в гл. 2

Под источником цифровой информации на рис. 1.1 понимается устройство, в котором выполнены все указанные выше операции. Такой источник создает сообщения, представляющие собой дискретные последовательности, образованные элементами, каждый из которых может принимать только конечное число различных значений.

Число возможных значений, которое может принять каждый элемент сообщения, называется размером алфавита источника. В простейшем и наиболее распространенном случае каждый элемент дискретного сообщения может принимать только одно из двух различных значений, т. е. имеем двоичный (бинарный) алфавит и сообщения, образованные с помощью такого алфавита, будут представлять последовательности, составленные при различном сочетании двух элементов x_1 и x_2 (например, нуля и единицы).

Различные дискретные (цифровые) сообщения отличаются друг от друга видом элементов, их числом и порядком следования. Каждое конкретное сообщение представляет собой одно из возможных сообщений, которые могут быть образованы данным источником. Если алфавит источника состоит из n элементов, а каждое создаваемое источником сообщение содержит m элементов, то общее число возможных сообщений равно

$$M = n^m. \tag{1.3}$$

Следовательно, каждое конкретное дискретное сообщение является одним из множества возможных. Поэтому передачу дискретных сообщений нужно рассматривать как случайный процесс, реализации которого представляют собой отдельные дискретные сообщения.

В дискретных сообщениях некоторые элементы могут появляться чаще, а некоторые — реже. Иными словами, вероятности появления тех или иных элементов в дискретном сообщении различны. Это обстоятельство учитывают с помощью ансамбля

$$x = \begin{pmatrix} x_1 x_2 \ldots x_i \ldots x_n \\ p_1 p_2 \ldots p_i \ldots p_n \end{pmatrix}, \tag{1.4}$$

который представляет собой совокупность элементов алфавита источника вместе с вероятностями их появления при образовании дискретного сообщения. Ансамбль

вида (1 4) часто называют статистической схемой дискретного источника сообщений.

Если совокупность элементов ансамбля (1.4) исчерпывающая, т е никаких других элементов при образовании сообщений не используется, то

$$\sum_{i=1}^{n} p_i = 1. \quad (1.5)$$

В дискретных сообщениях может иметь место не только преимущественное появление одних элементов по сравнению с другими, но и преимущественное появление одних сочетаний элементов по сравнению с другими. Иными словами, между отдельными группами элементов дискретного сообщения могут существовать статистические связи. Поэтому для более полного описания статистических свойств дискретных источников и создаваемых ими сообщений желательно иметь сведения о вероятности появления всех возможных сочетаний из двух, трех, четырех и т д элементов. Указанные сведения характеризуют степень нашего знания статистики дискретных сообщений и называются априорными сведениями Эти сведения могут быть получены из экспериментального исследования статистических свойств передаваемых сообщений, проведенного для данного класса сообщений заранее, или выполняемого в процессе передачи специальными устройствами. В последнем случае (почти неосуществимом практически) неизбежной оказывается длительная задержка в передаче, поскольку для получения достоверных априорных сведений сначала необходимо исследовать весьма продолжительные дискретные последовательности, а затем уже осуществить их передачу с учетом полученных сведений

Таким образом, на выходе источника дискретных сообщений имеем процесс, представляющий собой случайную дискретную последовательность, выраженную в цифровой форме Обозначим эту последовательность $\{x_i\}$.

Для более экономного использования линии связи, а также для уменьшения влияния различных помех и искажений передаваемая дискретная последовательность в общем случае может быть подвергнута дальнейшему преобразованию с помощью кодирующего устройства (кодера) Это преобразование может состоять из ряда операций, включающих учет статистики поступающей

случайной последовательности для устранения избыточной информации (статистическое кодирование), а также введение в дискретную последовательность дополнительных элементов для уменьшения влияния помех и искажений (помехоустойчивое кодирование).

Кроме того, в состав кодирующего устройства в общем случае могут входить устройства, осуществляющие статистический анализ поступающей дискретной последовательности если ее статистика заранее неизвестна, а также устройства, преобразующие дискретную последовательность из одного алфавита в другой. Естественно, что какие бы сложные операции ни выполнялись в таком кодирующем устройстве, они не должны изменять содержание исходных сообщений. Кодирование в таком широком смысле в цифровых системах еще не находит широкого применения в основном из-за значительных технических трудностей. По сути дела, такое кодирующее устройство должно представлять собой сложную специализированную электронно-вычислительную машину, работающую по некоторой заранее установленной программе.

Однако отдельные операции, о которых шла речь выше, в ряде систем уже успешно применяются. Несмотря на технические трудности, помехоустойчивое кодирование находит применение в некоторых системах радиоуправления, в системах дальней космической связи и ряде других систем, где существенное значение имеют вопросы повышения помехоустойчивости передачи информации.

В результате ряда преобразований на выходе кодера образуется дискретная случайная последовательность элементов, которую обозначим $\{x_{\text{к}i}\}$. Эта последовательность с помощью модулятора изменяет один или несколько параметров сигнала-переносчика информации, который генерируется в передающем устройстве. Изменения параметров сигнала-переносчика происходят в соответствии с изменениями дискретной последовательности и также дискретны. Такой процесс называют манипуляцией сигнала (в отличие от модуляции сигнала, когда его параметры изменяются непрерывно). Обычно осуществляется манипуляция только одного из параметров сигнала-переносчика. В результате манипуляции образуется случайная последовательность сигналов $\{s_{\text{к}i}(t)\}$, однозначно соответствующая дискретной

случайной последовательности $\{x_{кi}\}$. Сигнал, манипулированный дискретным сообщением, излучается в среду и выполняет роль переносчика информации между пунктами передачи и приема.

Таким образом, передающая часть цифровой линии связи осуществляет ряд операций, связанных с преобразованием дискретного сообщения в сигнал. Эти операции включают кодирование передаваемой информации (понимаемое в самом широком смысле) и манипуляцию сигнала.

Совокупность операций, связанных с преобразованием передаваемых сообщений в сигнал, называется способом передачи. Способ передачи можно описать операторным соотношением

$$\{s_{кi}(t)\} = D_{пер}\{x_i\} = D_м D_к\{x_i\}, \tag{1.6}$$

где $D_{пер}$ — оператор способа передачи; $D_к$ — оператор кодирования; $D_м$ — оператор манипуляции *.

Выражение (1.6) определяет математические преобразования, адекватные физическим операциям преобразования сообщений в сигналы.

В общем случае сигналы, излученные в среду распространения, претерпевают в ней ослабление (вследствие поглощения и рассеивания энергии) и искажения (замирания, отражения от неоднородностей среды, препятствий и т. п.). Поэтому сигналы $\{\hat{s}_{кi}(t)\}$, приходящие в точку приема, могут существенно отличаться от излучаемых передатчиком не только своей энергией, но и статистическими характеристиками и спектральным составом.

Влияние среды на распространяющиеся в ней сигналы можно также описать операторным соотношением

$$\{\hat{s}_{кi}(t)\} = D_с\{s_{кi}(t)\} = D_с D_{пер}\{x_i\}, \tag{1.7}$$

где $D_с$ — оператор среды распространения.

Учет влияния среды во многих случаях необходим; особенно это относится к средам, параметры которых непрерывно и случайным образом заметно меняются за время сеанса связи.

---

* Оператором называется символ, который ставится впереди функции и показывает выполнение некоторого действия, приводящего к новой функции

Кроме сигналов прошедших среду, на вход приемника цифровой линии связи попадают и различные аддитивные помехи внешнего происхождения (сигналы других систем, промышленные и атмосферные помехи, шумы среды и т. п.). К аддитивным помехам, действующим на входе, обычно относят также внутренние флюктуационные шумы приемника.

Таким образом, на входе приемника действует непрерывный случайный процесс

$$y(t) = \{\widehat{s}_{кi}(t)\} + \sum_{j=1}^{N} n_j(t), \qquad (1.8)$$

где $n_j(t)$ — случайный процесс, соответствующий одной из аддитивных помех; N — число независимых источников помех.

В приемной части цифровой линии связи принятая смесь сигнала и помех должна быть преобразована в дискретное сообщение, соответствующее переданному сообщению. Это преобразование состоит из ряда операций, обратных операциям, выполняемым в передающей части цифровой линии связи.

После фильтрации и усиления в избирательной части приемника сигнал демодулируется и на выходе демодулятора образуется дискретная последовательность $\{\hat{x}_{кi}\}$, которая должна соответствовать последовательности $\{x_{кi}\}$ на выходе кодирующего устройства. Однако из-за действия помех, влияния среды, погрешностей различных преобразований, вызванных неидеальностью и нестабильностью характеристик отдельных узлов и элементов системы, полного соответствия между $\{\hat{x}_{кi}\}$ и $\{x_{кi}\}$ может и не быть. Последовательность $\{\hat{x}_{кi}\}$ вводится в декодирующее устройство, которое выполняет операции по ее преобразованию в последовательность $\{\hat{x}_i\}$, соответствующую переданной цифровой последовательности $\{x_i\}$. Полнота этого соответствия зависит от ряда факторов: корректирующих возможностей кодированной последовательности $\{\hat{x}_{кi}\}$, уровня сигнала и помех и их статистики, свойств декодирующего устройства и т. п. Сформированная в результате декодирования цифровая последовательность $\{\hat{x}_i\}$ поступает к получателю информации.

Таким образом, приемная часть цифровой линии связи осуществляет ряд операций, связанных с преобразо-

ванием сигналов, принимаемых на фоне помех, в дискретные сообщения, выраженные в цифровой форме. Совокупность этих операций называется способом приема. Математически его можно описать операторным соотношением

$$\{\hat{x}_i\} = D_{\text{пр}}[y(t)] = D_{\text{пр}}\left[\{\hat{s}_{\text{к}i}(t)\} + \sum_{j=1}^{N} n_j(t)\right], \qquad (1.9)$$

где $D_{\text{пр}}$ — оператор способа приема. Его можно представить в виде произведения двух операторов, один из которых $D_{\text{дм}}$ соответствует процессу демодуляции сигналов, а второй $D_{\text{дк}}$ — декодированию. Тогда

$$\{\hat{x}_i\} = D_{\text{дк}} D_{\text{дм}}\left[\{\hat{s}_{\text{к}i}(t)\} + \sum_{j=1}^{N} n_j(t)\right]. \qquad (1.10)$$

Учитывая операторные соотношения, соответствующие передаче, влиянию среды и приему, весь процесс передачи информации по цифровым линиям связи можно также записать в операторной форме:

$$\{\hat{x}_i\} = D_{\text{пр}}\left[D_{\text{с}} D_{\text{пер}}\{x_i\} + \sum_{j=1}^{N} n_j(t)\right], \qquad (1.11)$$

которая с учетом смысла отдельных операторов может быть представлена в виде

$$\{\hat{x}_i\} = D_{\text{дк}} D_{\text{дм}}\left[D_{\text{с}} D_{\text{м}} D_{\text{к}}\{x_i\} + \sum_{j=1}^{N} n_j(t)\right]. \qquad (1.12)$$

Рассмотренный подход позволяет описать процесс передачи информации некоторой математической моделью, что открывает возможность четко поставить ряд важных проблем теории связи и наметить пути их решения. Одной из важнейших в этом ряду является проблема оптимизации линий связи.

1.5. ОСНОВНЫЕ КРИТЕРИИ И ПОКАЗАТЕЛИ КАЧЕСТВА РАБОТЫ ЦИФРОВЫХ СИСТЕМ ПЕРЕДАЧИ ИНФОРМАЦИИ

Из рассмотрения, проведенного в предыдущем параграфе, следует, что задача цифровой линии связи состоит в достаточно хорошем «приближении» (в некотором определенном смысле) принятого множества $\{\hat{x}_i\}$ к переданному множеству $\{x_i\}$. Иными словами, качество принимаемой информации зависит от степени указанного приближения.

Для количественной оценки этого приближения пользуются вероятностью ошибки приема отдельного элемента дискретного сообщения (см. § 1.2) или вероятностью ошибки приема всего дискретного сообщения.

Передача информации с требуемым качеством еще не дает оснований для суждения о том, хороша или плоха данная линия связи. Всегда желательно, чтобы линия связи обеспечивала передачу информации с требуемым качеством наиболее «экономно», т. е. с наименьшими «затратами», понимаемыми в некотором определенном смысле. Под такими затратами можно понимать затраты мощности, полосы частот, стоимости оборудования, его веса и т. д. и т. п., которые необходимы для передачи определенного количества информации (например, одной двоичной единицы).

Для количественной оценки затрат при передаче информации выбирают тот или иной критерий. Выбор критерия определяется рядом соображений и, как правило, представляет собой компромисс между стремлением учесть ряд особенностей работы линий связи и стремлением к простоте и удобству пользования. При выборе критерия необходимо учитывать требования и ограничения, которые наложены на некоторые характеристики и параметры системы связи.

В настоящее время в некоторых цифровых системах ограничивающим фактором первостепенной важности является величина излучаемой средней или пиковой мощности (например, в системах дальней космической связи), а в некоторых — используемая полоса частот (например, в системах передачи данных по стандартным телефонным каналам). Имеются системы, в которых не-

обходимо учитывать не один, а ряд ограничивающих факторов (вес, габариты, излучаемая мощность). Выбрать критерий, который был бы универсальным для любых систем передачи цифровой информации, трудно. Для одного класса систем лучше подходит один критерий, для другого — другой.

Наиболее широко в настоящее время в теории связи и в прикладных задачах применяются критерии, в соответствии с которыми системы связи оцениваются величиной затрат на передачу единицы количества информации при заданном качестве ее приема. Такие критерии можно назвать критериями удельных затрат. В соответствии с этими критериями лучшей оказывается та система, которая характеризуется меньшими удельными затратами.

Иногда более удобны критерии, в соответствии с которыми системы связи оцениваются количеством информации, приходящейся на единицу условных затрат при заданном качестве приема информации. Такие критерии можно назвать критериями удельной информации. В соответствии с этими критериями лучшей оказывается та система, которая характеризуется более высокой удельной информацией.

Нетрудно видеть, что критерии удельных затрат и удельной информации обратны по отношению друг к другу.

В дальнейшем при оценке качества работы различных систем передачи цифровой информации и сравнении этих систем между собою будем пользоваться критериями удельных затрат. Чтобы не усложнять задачу, из всех видов затрат будем учитывать только затраты энергии и полосы частот на передачу одной двоичной единицы информации (одного бита).

В соответствии с этим удельные энергетические затраты будем характеризовать величиной

$$\beta_E = E_0/N_0, \qquad (1.13)$$

где E_0 — энергия сигнала на входе приемника, соответствующая передаче одного бита информации с заданной достоверностью; N_0 — спектральная плотность мощности помехи на входе приемника.

Удельные затраты полосы будем характеризовать величиной

$$\beta_{\Delta f} = \Delta f_э/R, \qquad (1.14)$$

где R — информационная скорость передачи (бит/с); $\Delta f_э$ — эквивалентная полоса пропускания приемного устройства, соответствующая ширине спектра используемого сигнала *.

Такие критерии впервые были внедрены Сандерсом [16] для оценки качества работы систем связи, в которых основной помехой являются флюктуационные шумы с равномерной спектральной плотностью N_0. Эти критерии оказались очень удобными и получили широкое распространение в теории связи и ее приложениях.

Если спектральную плотность нельзя считать равномерной (помеха типа «окрашенного» шума), можно приближенно пользоваться усредненной величиной удельных затрат энергии

$$\overline{\beta_E} = E_0/\overline{N_0(f)}, \tag{1.15}$$

где $\overline{N_0(f)}$ — средняя величина спектральной плотности помех в полосе пропускания $\Delta f_э$, определяемая выражением

$$\overline{N_0(f)} = \frac{1}{\Delta f_э} \int\limits_0^{\Delta f_э} N_0(f)\, df. \tag{1.16}$$

Значения β_E и $\beta_{\Delta f}$ можно рассматривать как обобщенные показатели качества работы системы связи, оцениваемой в соответствии с критериями затрат.

Как уже указывалось ранее (см. § 1.2), качество передачи цифровой информации оценивается достоверностью и скоростью ее передачи. При оценке же качества работы системы передачи информации наряду с этими абсолютными показателями необходимо также учитывать обобщенные показатели, рассмотренные выше.

Итак, качество работы системы передачи цифровой информации можно характеризовать следующими показателями:

1) вероятностью ошибки при передаче одного элемента дискретного сообщения;

---

* Как известно из курса радиоприемных устройств, полоса пропускания приемника определяется в основном полосой пропускания его усилителя промежуточной частоты.

2) информационной и технической скоростью передачи сообщений;

3) энергетическими затратами на передачу одной двоичной единицы информации (удельные затраты энергии);

4) затратами полосы на передачу одной двоичной единицы информации (удельные затраты полосы).

Необходимо отметить, что перечисленные показатели не затрагивают такие важные стороны, связанные с технической реализацией систем, как стоимость, вес, габариты и т. д. Учет этих факторов представляет большую самостоятельную проблему, относящуюся к области проектирования и конструирования систем. В данной книге эти вопросы не рассматриваются.

1.6. ПРОБЛЕМА ОПТИМИЗАЦИИ ЦИФРОВЫХ ЛИНИЙ СВЯЗИ

Если критерии оценки качества работы линии связи выбраны, то ее показатели в соответствии с этими критериями будут тем выше, чем лучше выбраны способы передачи и приема. Можно представить такую линию связи, в которой при равных прочих условиях соответствующие операции приема и передачи выполнены «наилучшим» образом и, следовательно, показатели качества работы линии связи самые высокие. Линия связи, в которой применены наилучшие способы передачи и приема (с точки зрения выбранных критериев), оптимальна или идеальна.

В любой линии связи имеются три фактора, определяющие степень соответствия принятых сообщений переданным (достоверность или точность передачи): 1) наличие помех; 2) неидеальность и нестабильность различных характеристик системы связи; 3) неидеальность характеристик среды распространения сигналов.

Следовательно, оптимизация линии связи с математической точки зрения сводится к такой оптимизации выражения (1.11), чтобы при требуемой достоверности передачи информации обеспечивался минимум затрат в соответствии с принятым критерием. Сказанное можно записать в виде формального соотношения

$$\{\hat{x}_i\} = \text{opt}\left\{D_{\text{пр}}\left[D_{\text{с}}D_{\text{пер}}\{x_i\} + \sum_{j=1}^{N} n_j(t)\right]\right\}, \qquad (1.17)$$

которое соответствует передаче информации с минимальными затратами при условии заданного соответствия между принятым множеством $\{\hat{x}_i\}$ и переданным множеством $\{x_i\}$.

В общем случае оператор $D_{\text{с}}$, определяющий влияние среды, может оказывать существенное влияние на передачу информации. Особенно это относится к средам с переменными случайными параметрами. Непосредственно воздействовать на оператор $D_{\text{с}}$ и изменить его в желаемую сторону в настоящее время не представляется возможным, так как свойства среды практически неуправляемы. По этой причине оптимизация линии связи может быть достигнута только оптимальным совместным выбором операторов $D_{\text{пер}}$ и $D_{\text{пр}}$, т. е. соответствующим совместным выбором оптимальных способов передачи и приема. При такой постановке задачи оптимизация линии связи сводится к оптимизации системы связи с учетом действия помех и влияния среды. Чем более полны и достоверны априорные сведения о помехах и среде, тем успешнее могла бы быть решена эта задача.

Несмотря на то, что задачу оптимизации линии связи можно сформулировать в достаточно общем виде, практическая ценность такой постановки невелика. Дело в том, что до сих пор нет достаточно эффективных методов оптимизации сложных систем в целом. Попытка четко сформулировать и решить такие задачи в весьма общем виде наталкиваются на большие математические трудности, преодолеть которые в настоящее время не удается.

Как известно [17], эффективная теория оптимизации систем связи должна давать ответ на следующие вопросы.

1. Какова структура оптимальной системы, т. е. какие конкретные операции необходимо выполнить, чтобы обеспечить оптимальную передачу и прием информации в заданных условиях и при определенных наложенных ограничениях?

2. Как оценить качество работы системы, т. е. как получить количественные характеристики оптимальной передачи сообщений?

3. Как сравнить оптимальные системы с реальными и каковы пути приближения реальных систем к оптимальным (реализация субоптимальных систем)?

Ответ на поставленные вопросы в настоящее время пытаются получить либо существенно упрощая рассматриваемые модели линий связи, либо вводя ряд ограничений и фиксированных условий. В соответствии с этим сложилось два направления развития теории оптимизации систем.

Первое связано с разработкой теории оптимизации всей системы связи. Оно основывается на рассмотрении упрощенных идеализированных моделей линий связи. Задачу оптимизации всей системы часто заменяют задачей оптимизации ее отдельных подсистем, т. е. проводят оптимизацию по частям в соответствии с некоторыми установленными критериями. Такой подход упрощает задачу и в некоторых случаях позволяет решить ее до конца. Однако необходимо иметь в виду, что оптимизация системы по частям не эквивалентна оптимизации всей системы в целом.

Второе направление связано с разработкой теории оптимальных систем связи при условии, что способы передачи заданы. Это направление получило название теории оптимальных методов приема.

Рассмотрим более детально суть указанных направлений.

1.6.1. Оптимизация системы связи в целом

Для этого направления характерен ряд существенных допущений, предположений и идеализаций, относящихся к среде распространения, к характеристикам передающей и приемной частей системы связи, а также к помехам. Остановимся на некоторых, часто применяемых в теории допущениях и идеализациях.

1. Допущение идеальности среды. Считается, что среда распространения сигналов однородна, ее свойства неизменны, в ней отсутствует поглощение и единственной причиной ослабления сигналов является сферичность расхождения излучаемых сигналов в пространстве. При прохождении через идеальную среду сигналы не искажаются и в зависимости от расстояния меняется только их масштаб по амплитуде. Оператор среды D_c в этом случае представляет собой коэффициент, зависящий только от расстояния между передатчиком и при-

емником. При фиксированном расстоянии (передатчик и приемник неподвижны) $D_c = \text{const}$.

2. Допущение идеальности и стабильности характеристик системы связи. Считается, что все характеристики приемной и передающей частей системы таковы, что в отсутствие помех вся переданная информация полностью и без искажений поступала бы к получателю.

3. Допущение о том, что единственным видом помех является шум с нормальным законом распределения.

Проблема оптимизации системы связи в целом впервые была поставлена и частично решена одним из основоположников теории информации американским ученым К. Шенноном [18]. Основываясь на перечисленных выше допущениях и считая, что основная задача системы связи состоит в максимизации средней скорости передачи сообщений по каналу связи, Шеннон доказал ряд теорем, утверждающих существование совместных оптимальных методов передачи и приема. Иными словами, он доказал принципиальную возможность оптимизации всей системы связи в целом по максимуму средней скорости передачи. Шеннону удалось получить результаты, которые дают возможность количественно оценивать качество работы оптимальной системы связи. Однако на остальные вопросы ответа получить не удалось. Вопрос о структуре оптимальной системы (иными словами, задача синтеза) остался открытым — теория Шеннона не указала конкретные пути ее решения.

Решение задачи синтеза оптимальной системы предполагает совместный учет всех операций, связанных с преобразованием сообщений в сигналы и обратным преобразованием смеси сигналов с помехами в сообщения. Эти операции ранее были названы способами передачи и приема. В теории информации их часто называют кодированием и декодированием, трактуя эти понятия очень широко и включая в них операции модуляции и демодуляции.

Таким образом, в соответствии с подходом Шеннона оптимальная система связи — это такая система, в которой применены «наилучшие» способы кодирования и декодирования информации, обеспечивающие максимум средней скорости передачи информации по каналу связи. Оптимизация в указанном смысле соответствует отысканию структуры системы, заключенной между сечениями A и B на рис. 1.1.

Несмотря на ряд существенных допущений (о которых шла речь выше), раскрыть суть оптимальных процедур кодирования и декодирования, а также указать пути их реализации (т. е. установить структуру оптимальной системы связи) до сих пор не удалось. Поэтому для решения задачи синтеза оптимальной системы связи приходится идти по пути дальнейших упрощений, в частности, рассматривать задачу, не учитывая специальные методы кодирования и декодирования, понимая под способами передачи и приема процедуры, связанные только с модуляцией и демодуляцией сигналов. В такой постановке оптимизация системы соответствует отысканию ее структуры, заключенной между сечениями C и D на рис. 1.1. Подобную задачу называют оптимизацией системы «модулятор — демодулятор». В результате решения этой задачи должен быть найден «наилучший» вид сигнала (т. е. установлен вид оптимального сигнала) и соответствующий ему оптимальный способ приема такого сигнала. Поэтому указанную выше задачу иногда также называют оптимизацией системы «сигнал — фильтр». (Решение такой задачи для частного случая бинарного канала выполнено в гл. 3.)

Возможна и другая упрощенная постановка задачи. Суть ее сводится к следующему. Часть системы, включая модулятор и демодулятор (часть между сечениями C и D на рис. 1.1), рассматривается как «черный ящик» и характеризуется стохастической матрицей трансформации множества $\{x_{\text{к}i}\}$ в множество $\{\hat{x}_{\text{к}i}\}$. Такая матрица определяет все возможные вероятности переходов элементов алфавита множества $\{x_{\text{к}i}\}$ в элементы алфавита множества $\{\hat{x}_{\text{к}i}\}$. Иными словами, элементами матрицы $\|p_{ij}\|$ являются переходные вероятности p_{ij}, характеризующие переход i-го элемента алфавита множества $\{x_{\text{к}i}\}$ в j-й элемент алфавита множества $\{\hat{x}_{\text{к}i}\}$.

В простейшем случае, когда алфавит входного и выходного множеств двоичный, матрица имеет вид

$$\|p_{ij}\| = \begin{Vmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{Vmatrix} \qquad (1.18)$$

и «черный ящик» характеризуется диаграммой переходов, показанной на рис. 1.2. Значения переходных вероятностей определяются энергией и свойствами сигналов, характеристиками демодулятора, видом и уровнем помех. Задавая матрицу переходных вероятностей, мы

формализуем задачу и отвлекаемся от конкретного содержания и свойств той части системы, которая представлена в виде «черного ящика». Одна и та же матрица может соответствовать различному физическому содержанию этого «черного ящика». Однако она позволяет достаточно полно описать статистические особенности этого ящика и в этом смысле является его важнейшей интегральной статистической характеристикой.

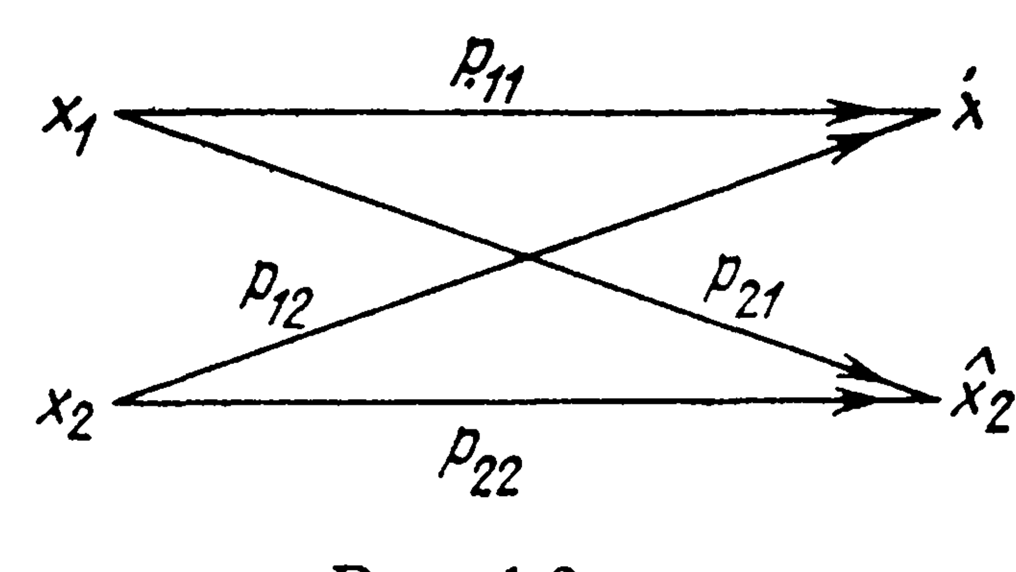


Рис. 1.2.

В данном случае оптимизация сводится к отысканию «наилучшей» структуры системы «кодер—декодер» при заданной статистической характеристике остальной части. Таким образом, можно пытаться оптимизировать систему связи, оптимизируя ее отдельные части: систему «сигнал — фильтр» и систему «кодер — декодер».

Необходимо отметить, что несмотря на многочисленные исследования [19, 20, 23, 25], затрагивающие различные стороны этой задачи, проблема оптимизации даже и в такой упрощенной постановке еще далека от завершения. Методы оптимизации радиоэлектронных устройств рассматриваются также в [26].

1.6.2. Оптимизация приемной части системы связи

Для такого направления характерен более узкий подход к проблеме оптимизации линии связи. Считается, что вся совокупность операций, определяющая способ передачи, задана заранее. Задача сводится к отысканию такого способа приема, который был бы оптимальным для заданного способа передачи и конкретных условий и ограничений, характерных для данной линии связи.

Сказанное можно записать в виде формального соотношения

$$\{\hat{x}_i\} = \text{opt}\, D_{\text{пр}}\,[y(t)] \tag{1.19}$$

или с учетом (1.9)

$$\{x_i\} = \text{opt}\, D_{\text{пр}} \left[\{\hat{s}_{\text{к}i}(t)\} + \sum_{j=1}^{N} n_j(t) \right], \tag{1.20}$$

при условии, что вид помех, их параметры и все операции $D_{пер}$, связанные со способом передачи, заданы заранее (известны), а влияние среды учтено тем, что исходная (излученная) совокупность сигналов $\{s_{кi}(t)\}$ преобразуется в отличающуюся от нее принимаемую совокупность $\{\hat{s}_{кi}(t)\}$.

Решение задачи обычно сопровождается рядом условий и ограничений, к числу которых могут относиться ограничения на пиковую или среднюю мощность излучаемых сигналов, вид аддитивных помех, законы распределения флюктуаций параметров сигнала в среде со случайными параметрами и т. д.

Рассматриваемое направление оптимизации является предметом теории оптимальных методов приема. Основы этой теории были разработаны советским ученым В. А. Котельниковым в 1946—1947 гг. и изложены им в классической работе «Теория потенциальной помехоустойчивости» [21]. Эта работа послужила толчком к интенсивным исследованиям различных вопросов оптимального приема.

Наиболее детально к настоящему времени разработаны вопросы оптимального приема в каналах с известной статистикой сигналов и помех.

Несмотря на ряд существенных допущений и ограничений, при которых обычно решаются различные задачи оптимального приема, получаемые результаты имеют большую ценность, так как позволяют ответить на многие принципиально важные вопросы. В частности, сравнивая показатели оптимального приема с показателями реального способа приема, можно судить о том, насколько целесообразно дальнейшее совершенствование этого способа. Располагая показателями оптимального приема при разных способах передачи, можно установить, какие из этих способов передачи наиболее целесообразны. Важным достоинством теории оптимальных методов приема является то, что во многих случаях она дает возможность не только найти показатели, но и установить структуру оптимального приемника.

Теория оптимальных методов приема является эффективным средством в руках разработчика новых систем передачи информации, позволяющим ему заранее разобраться, какой из возможных систем и в каком случае необходимо отдать предпочтение; можно ли повысить качественные показатели системы, совершенствуя

только способ приема; каковы пути реализации способа приема, близкого к оптимальному, и т. п.

Сочетание общего и частичного подхода к оптимизации систем связи позволяет более обоснованно и с меньшими затратами времени и средств решать задачи проектирования и разработки таких систем.

Вместе с тем важно отметить, что любая теория оптимизации систем всегда базируется на ряде идеализаций и упрощений действительной картины. Эту мысль очень хорошо выразил один из основоположников общей теории связи американский ученый Д. Миддлтон [17]: «В попытках моделирования действительности упрощение, приближенное рассмотрение и компромисс являются скорее правилом, чем исключением, если желательно получить количественные соотношения». Поэтому необходимо правильно оценивать возможности и результаты теории оптимальных систем — не преувеличивая и не преуменьшая их значения. Эта теория, как правило, указывает направления поисков и предельные показатели качества работы систем, но не дает готовых рецептов, как осуществить необходимую систему.

Создание реальных систем является сложным процессом, опирающимся на хорошее знание теории, инженерный опыт и смекалку. Важная роль результатов теории оптимизации в этом процессе будет показана при рассмотрении ряда вопросов (см. § 1.7; гл. 3 и 6).

Помимо задач оптимизации систем, в инженерных приложениях возникает и другая весьма важная задача — эффективное использование уже действующих линий и каналов связи, так как в их создание вложены большие средства. Это вызывает значительное развитие методов защиты передаваемой информации от помех путем помехоустойчивого кодирования; методов устранения избыточности в передаваемых сообщениях; методов понижения требуемого отношения сигнал/шум при сохранении заданной точности передачи информации; совершенствование методов уплотнения и разделения каналов при многоканальной передаче; совершенствование методов и способов модуляции и демодуляции и т. д.

Итак, развитие современных систем передачи цифровой информации идет по двум направлениям: 1) отыскание новых методов и способов передачи и приема информации, приближающих показатели работы систем

к предельно возможным; 2) совершенствование существующих систем с целью их более эффективного использования.

1.7. ПРЕДЕЛЬНЫЕ ВОЗМОЖНОСТИ СИСТЕМ ПЕРЕДАЧИ ЦИФРОВОЙ ИНФОРМАЦИИ. ВЫБОР ВИДА СИГНАЛОВ

В предыдущем параграфе указывалось, что с учетом ряда допущений и идеализаций К. Шеннону удалось доказать ряд теорем об оптимальных каналах связи. В частности, для гауссовского канала с белым шумом и ограниченной средней мощностью сигнала Шеннон получил широко известную формулу, определяющую максимум средней скорости передачи информации по такому каналу (пропускную способность канала) [18]:

$$\max R = C = \Delta f_э \log_2 \left(1 + \frac{P_c}{P_ш}\right) = \Delta f_э \log_2 \left(1 + \frac{P_c}{N_0 \Delta f_э}\right), \tag{1.21}$$

где C — пропускная способность канала; $\Delta f_э$ — ширина полосы частот, занимаемой каналом (ширина спектра сигнала); P_c — средняя мощность сигнала; $P_ш$ — средняя мощность шума в канале; N_0 — спектральная плотность белого гауссовского шума.

Эта формула характеризует предельные возможности канала, в котором применены «наилучшие» способы передачи и приема (оптимальное кодирование и декодирование), обеспечивающие согласование производительности источника информации с пропускной способностью канала. При этом вся информация, создаваемая источником, полностью поступает к получателю, т. е. потери информации в канале за счет действия помех не происходит. Это означает, что между переданным и принятым сообщением имеется полное соответствие и ошибки отсутствуют.

Необходимо отметить, что формула (1.21) справедлива только при физически нереализуемых условиях: бесконечном времени передачи информации (бесконечное время кодирования приводит к бесконечному времени задержки поступления информации к потребителю) и неизменных условиях работы канала. Сказанное озна-

чает, что формулу (1.21) нужно рассматривать как асимптотическое соотношение, определяющее физически недостижимую границу скорости передачи информации *.

В каналах, где кодирование неоптимально (или отсутствует вовсе), скорость передачи меньше пропускной способности канала:

$$R < \max R = C,$$

а ошибка при передаче информации не равна нулю.

Несмотря на существенные отличия условий работы реальных каналов связи от условий работы оптимальных в смысле Шеннона каналов, очень интересно выяснить предельные возможности оптимальных каналов и установить границы, которые принципиально не могут быть превзойдены ни при каких способах передачи и приема.

В соответствии с критериями, рассмотренными в § 1.5, будем характеризовать затраты энергии и полосы на передачу одной двоичной единицы информации в оптимальном (идеальном) канале величинами

$$\beta_E = E_0/N_0 = P_c T_0/N_0, \tag{1.22}$$

$$\beta_{\Delta f} = \Delta f_э/R_{max} = \Delta f_э/C, \tag{1.23}$$

где E_0 и T_0 — энергия сигнала и время, соответствующие передаче одной двоичной единицы информации в идеальном канале.

Приняв во внимание, что по определению

$$\max R T_0 = C T_0 = 1 \text{ дв. ед.} \tag{1.24}$$

и учитывая выражения (1.22) и (1.23), формулу (1.21) нетрудно преобразовать к виду

$$\frac{\Delta f_э}{C} \log_2 \left(1 + \frac{E_0}{N_0} \frac{C}{\Delta f_э}\right) = 1 \tag{1.25}$$

или

$$\log_2 \left(1 + \frac{\beta_E}{\beta_{\Delta f}}\right)^{\beta_{\Delta f}} = 1. \tag{1.26}$$

---

* В более поздних работах Шеннон, а затем и ряд других ученых установили границы пропускной способности с учетом конечного (хотя и продолжительного) времени передачи информации (см., например, [22, 23]). При этом ошибка передачи информации уже не может быть полностью устранена и имеет некоторое конечное значение. Эти вопросы выходят за рамки данной книги, и мы на них останавливаться не будем.

После несложных преобразований последнее выражение можно привести к виду

$$\beta_E = \beta_{\Delta f}(2^{1/\beta_{\Delta f}} - 1). \qquad (1.27)$$

Выражение (1.27) определяет функциональную связь между удельными затратами энергии и полосы в идеальном гауссовском канале связи. Нетрудно показать,

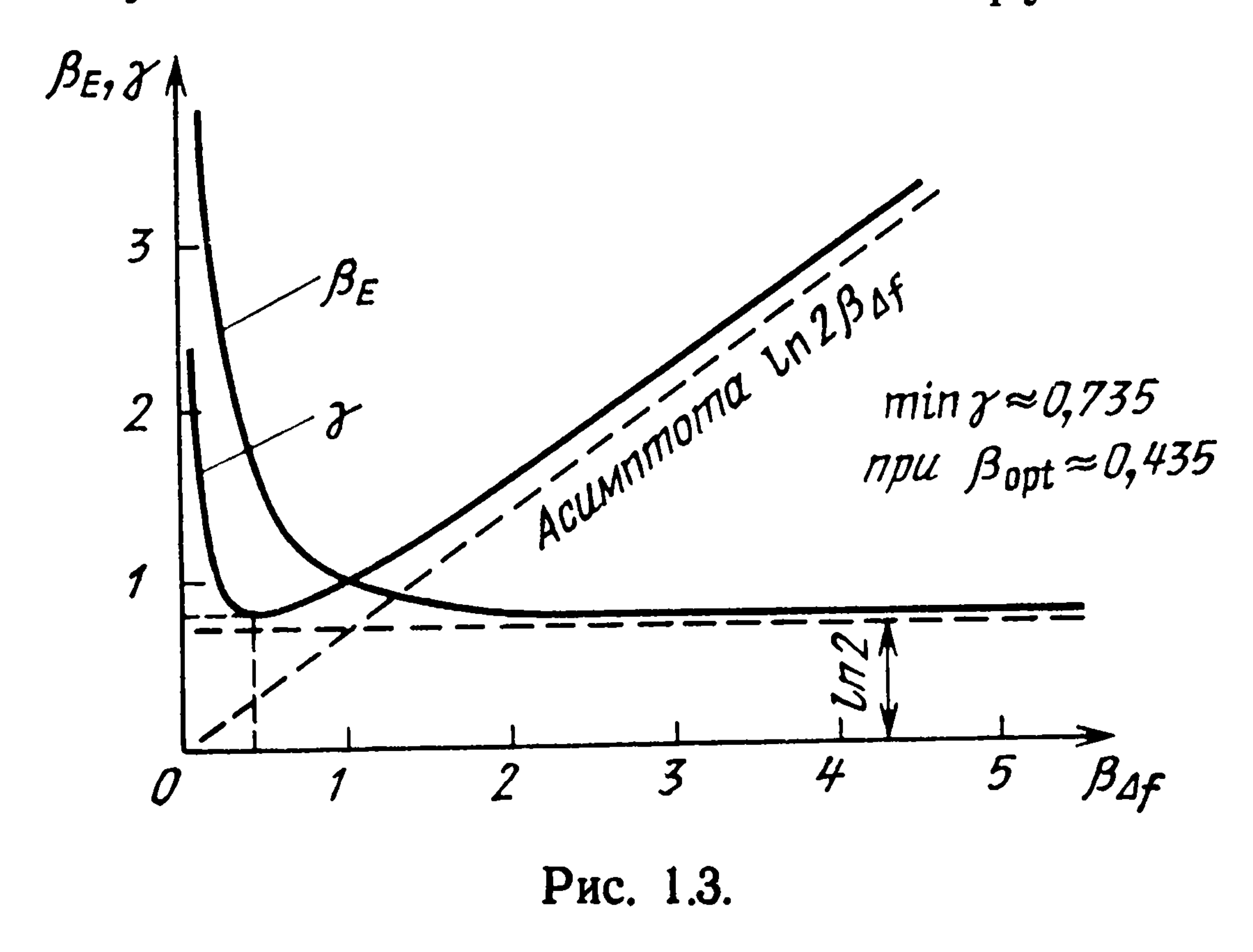


Рис. 1.3.

что при увеличении удельных затрат полосы удельные затраты энергии в идеальном канале уменьшаются, стремясь в пределе к величине

$$\lim_{\beta_{\Delta f}\to\infty} \beta_E = \min \beta_E = 1/\log_2 e = \ln 2 \approx 0{,}7. \qquad (1.28)$$

Зависимость (1.27) показана на рис. 1.3. Эта зависимость называется границей Шеннона для идеального гауссовского канала связи. Она представляет собой геометрическое место точек, координаты каждой из которых соответствуют показателям оптимальной (идеальной) системы в смысле Шеннона.

Из проведенного рассмотрения следует, что в соответствии с теорией Шеннона возможно бесконечное множество оптимальных систем, каждая из которых характеризуется своими показателями удельных затрат β_E и $\beta_{\Delta f}$.

График на рис. 1.3 показывает, что желание улучшить один из показателей приводит к неизбежному ухудшению другого. Оптимальные системы с малыми затратами энергии требуют значительных затрат полосы

и, наоборот, системы с малыми затратами полосы — значительных затрат энергии. Так как показатели реальных систем не могут быть выше показателей оптимальных систем, то принципиально реализуемым системам связи соответствуют только те точки плоскости (β_E, $\beta_{\Delta f}$), которые лежат в области выше границы Шеннона.

Выводы, вытекающие из теории Шеннона, имеют принципиальное значение, так как указывают пути решения ряда важных задач прикладного характера. К их числу относятся задачи выбора вида сигнала. Их важность объясняется тем, что показатели качества работы проектируемой информационной системы существенно зависят от того, насколько правильно выбран вид сигнала с учетом специфики области применения, конкретных требований к данной системе, а также ограничений, наложенных на некоторые ее характеристики и параметры. Указанные обстоятельства делают задачу выбора вида сигнала весьма сложной. Естественно, что при решении этой задачи приходится прибегать к некоторым упрощениям и допущениям, и поэтому нет полной уверенности, что найденные сигналы близки к оптимальным для данной информационной системы.

Отмеченные трудности можно уменьшить, если рассмотреть проблему выбора сигналов в идеальных каналах связи и установить особенности различных возможных классов сигналов в таких каналах. Очевидно, что в реальных каналах класс возможных сигналов не может быть шире, чем в идеальных. Следовательно, открываются пути отыскания таких видов сигналов, которые в данных конкретных условиях и при некоторых наложенных ограничениях окажутся наиболее приемлемыми.

Рассмотрим затронутую выше проблему подробнее. Из общей теории информации известно, что максимальная средняя производительность H_t стационарного источника дискретной информации определяется величиной [18]

$$\max H_t = \lim_{T\to\infty} (\log_2 M/T), \tag{1.29}$$

где M — число различных сообщений длительностью T, вырабатываемых дискретным источником.

В наиболее интересном с точки зрения практики частном случае, когда источник, располагающий алфавитом из n символов, вырабатывает сообщения, каждое

из которых состоит из m символов одинаковой длительности τ_0, выражение (1.29) принимает вид

$$\max H_t = \lim_{m\to\infty} (\log_2 n^m/m\tau_0) = \log_2 n/\tau_0. \qquad (1.30)$$

Это выражение показывает, что максимальная производительность источника дискретной информации в данном частном случае определяется только основанием алфавита n и длительностью информационного символа τ_0.

В оптимальном (идеальном) по Шеннону канале связи источник информации всегда согласован с каналом, т. е. его производительность равна пропускной способности канала:

$$\max H_t = \max R = C. \qquad (1.31)$$

Учитывая это условие и выражение (1.24), имеем

$$\log_2 n/\tau_0 = C = 1/T_0, \qquad (1.32)$$

отсюда следует

$$T_0 = \tau_0/\log_2 n. \qquad (1.33)$$

Последнее выражение устанавливает связь между временем T_0, затрачиваемым на передачу одной двоичной единицы информации, и длительностью τ_0 информационного символа. Полученное выражение позволяет записать удельные затраты полосы в идеальном канале в виде

$$\beta_{\Delta f} = \Delta f_э/\max R = \Delta f_э T_0 = \Delta f_э \tau_0/\log_2 n = Б_с/\log_2 n. \qquad (1.34)$$

Как известно, величина

$$Б_с = \Delta f_э \tau_0 \qquad (1.35)$$

называется базой сигнала. Она определяет соотношение между шириной спектра используемого сигнала и длительностью информационного символа. Из теории сигналов известно, что минимальное значение этой величины приблизительно равно единице [24].

Сигналы с базой, равной единице (или близкие к этому), называются простыми, а с базой, значительно превышающей единицу, — сложными. Иногда в литературе сложные сигналы называют широкополосными. Этот термин, на наш взгляд, не удачен, так как он отражает только то обстоятельство, что при большом значении базы соблюдается условие $\Delta f_э \gg 1/\tau_0$. Это вовсе не

означает, что абсолютное значение полосы, занимаемое спектром сигнала, велико. Одну и ту же полосу часто могут занимать как сложные, так и простые сигналы. Различие в сигналах определяется не полосой частот, которую они занимают, а их свойствами *.

Рассмотрим два характерных случая.

Малые затраты полосы ($\beta_{\Delta f} \leqslant 1$). В этом случае в соответствии с формулой (1.34) используемые сигналы характеризуются соотношением

$$Б_с \leqslant \log_2 n, \tag{1.36}$$

откуда следует условие

$$n \geqslant 2^{Б_с}. \tag{1.37}$$

Условие (1.37) означает, что если применять сложные сигналы с большой базой ($Б_с \gg 1$), то необходимо использовать огромные основания алфавита. Очевидно, что в данном случае это практически неприемлемый путь решения задачи.

Если применять сигналы с минимально возможной базой ($Б_с = 1$), то в рассматриваемом случае

$$n = 2^{1/\beta_{\Delta f}} \geqslant 2. \tag{1.38}$$

Из этого выражения следует, что для получения малых затрат полосы необходимо выбирать сигналы с минимально возможной базой и использовать алфавиты с основанием больше двух. Сигналы, удовлетворяющие таким условиям, называются многоосновными (многоалфавитными). К таким сигналам, в частности, относятся сигналы с многократной фазовой манипуляцией.

Естественно, что, стремясь выбрать сигналы с малыми затратами полосы, нельзя забывать о том, что при этом могут потребоваться значительные затраты энергии. Представление о величине затрат в идеальной системе для простых сигналов с различными основаниями можно получить из табл. 1.1, рассчитанной по формулам (1.34) и (1.27).

В этой таблице величина $1/\beta_{\Delta f}$ характеризует удельную скорость передачи (число бит в секунду, приходящихся на один герц полосы). Из рассмотрения данных табл. 1.1

---

* Особенности и свойства сложных сигналов рассмотрены в гл. 6.

Таблица 1.1

n	2	4	8	16	32	64
$\beta_{\Delta f}$	1	0,5	0,33	0,25	0,2	0,167
$1/\beta_{\Delta f}$	1	2	3	4	5	6
β_E	1	1,5	2,31	3,75	6,2	10,5

следует, что увеличение основания алфавита приводит к сравнительно малому уменьшению затрат полосы и к заметному увеличению затрат энергии. Поэтому выбор сигналов с большими основаниями должен осуществляться очень осторожно. Обычно сигналы с основаниями $n>8$ не применяют по ряду причин (некоторые из них подробно рассматриваются в гл. 6).

Сигналы рассмотренного класса находят широкое применение в системах передачи данных, использующих радиотелефонные и другие каналы с ограниченной полосой пропускания.

Малые затраты энергии ($\beta_E \leqslant 1$). Выше было установлено (см. формулу (1.27) и вытекающие из нее результаты), что стремление к минимально возможным затратам энергии даже в идеальном канале связи приводит к очень значительным затратам полосы. Следовательно, условию ($\beta_E \to \min \beta_E$) соответствует условие ($\beta_{\Delta f} \to \infty$). В этом случае используемые сигналы характеризуются соотношением

$$Б_с \gg \log_2 n. \qquad (1.39)$$

Так как минимальное основание алфавита (а следовательно, и основание сигнала) равно двум, то для получения энергетических затрат, приближающихся к минимально возможным, необходимо применять сложные сигналы с большой базой. Однако расчеты показывают, что условие (1.39) слишком жесткое, так как энергетические затраты очень близки к минимально возможным уже при выполнении условия

$$Б_с \geqslant 10 \log_2 n.$$

Если $n=2$, то уже при базе $Б_с=10$ величина β_E практически совпадает с минимально возможным значением затрат энергии.

Возникает вопрос, какой же смысл в том, чтобы применять сигналы с большой базой, если увеличение базы сигнала практически не снижает энергетические затраты и только увеличивает затраты полосы? Дело в том, что сложные сигналы обладают рядом важных свойств, которые особенно ярко проявляются при больших значениях базы. Одно из таких свойств состоит в том, что прием подобных сигналов принципиально возможен при очень малом отношении сигнала к шуму на входе приемника. Действительно, если представить выражение для удельных затрат энергии в виде

$$\beta_E=\frac{E_0}{N_0}=\frac{P_c T_0 \Delta f_э}{N_0 \Delta f_э}=\frac{P_c}{P_ш}\frac{Б_c}{\log_2 n} \qquad (1.40)$$

и учесть соотношения (1.27) и (1.34), то необходимое отношение сигнала к шуму в идеальном канале можно выразить формулой

$$\frac{P_c}{P_ш}=\frac{\beta_E}{\beta_{\Delta f}}=2^{1/\beta_{\Delta f}}-1=\frac{\beta_E}{Б_c}\log_2 n. \qquad (1.41)$$

Из этого выражения следует, что при соответствующем выборе базы сигнала величина необходимого отношения сигнал/шум может быть сколь угодно малой. Это означает, что принимаемый сигнал будет «замаскирован» шумами. Отмеченное свойство сложных сигналов с большой базой открывает принципиальные возможности построения таких систем связи, работу которых трудно обнаружить. Несмотря на большие трудности технической реализации информационные системы со сложными сигналами находят применение в области управления, в космической радиосвязи и т. п.

Выше были рассмотрены некоторые вопросы выбора вида сигнала в идеальных каналах связи для двух крайних случаев: 1) когда желательно иметь малые затраты полосы и 2) когда желательно иметь малые затраты энергии при передаче информации. Однако имеется ряд систем (в частности, некоторые космические системы, радио- и гидротелеметрические системы) при реализации которых существенное значение имеют оба указанных фактора. В таких случаях желательно выбрать

класс сигналов, который обеспечивал бы сравнительно малые затраты и полосы и энергии.

Рассмотрим эту задачу применительно к идеальному каналу связи. Чтобы учесть оба вида затрат, введем обобщенный показатель качества (обобщенные удельные затраты)

$$\gamma_{E,\ \Delta f} = \beta_{\Delta f} \beta_E. \quad (1.42)$$

С учетом выражения (1.27) для идеальной системы этот показатель принимает вид

$$\gamma_{E,\ \Delta f} = \beta^2_{\Delta f} (2^{1/\beta_{\Delta f}} - 1). \quad (1.42)$$

График зависимости обобщенных затрат от величины $\beta_{\Delta f}$ приведен на рис. 1.3. Он представляет собой геометрическое место точек, координаты каждой из которых соответствуют показателям оптимальной системы.

Кривая разбивает всю плоскость ($\gamma_{E,\ \Delta f}$, $\beta_{\Delta f}$) на две части. Все точки, лежащие выше кривой, относятся к принципиально реализуемым системам. Из рассмотрения графика следует, что при некоторой величине $\beta_{\Delta f}$ обобщенные затраты в оптимальной системе достигают минимального значения. Расчеты показывают, что $\min \gamma_{E,\ \Delta f} \approx 0{,}735$ при $\beta_{\Delta f} \approx 0{,}435$.

В области $\beta_{\Delta f} \approx 0{,}35 - 0{,}5$ величина обобщенных затрат мало отличается от минимального значения. Если учесть выражение (1.34), то нетрудно видеть, что указанная область соответствует классу простых сигналов с основаниями от четырех до восьми (n=4—8). Если ограничить обобщенные затраты величиной $\gamma_{E,\ \Delta f} \leqslant 1$, то область возможных простых сигналов расширится от $n=2$ до $n=16$, т. е. охватит весь класс простых многоосновных сигналов, представляющих практический интерес.

Проведенное рассмотрение возможных дискретных сигналов в идеальных по Шеннону каналах связи может служить хорошим ориентиром при решении ряда практических задач по выбору наиболее подходящего для заданных требований класса сигналов. Дальнейшая оптимизация информационной системы может быть сведена к выбору конкретного вида сигналов в пределах наиболее подходящего класса таких сигналов. Некото-

рые из затронутых вопросов будут более подробно рассмотрены в гл. 6.

В заключение отметим, что показатели качества работы лучших современных цифровых систем связи еще заметно уступают показателям оптимальных систем, особенно по удельным затратам энергии. Так например, в идеальных оптимальных системах удельным затратам полосы $\beta_{\Delta f} \approx 0,3-0,5$ соответствуют удельные затраты энергии $\beta_E \approx 2,5-1,5$. В лучших реальных системах при тех же затратах полосы энергетические затраты возрастают в десятки и даже сотни раз по сравнению с идеальными. При этом необходимо иметь в виду, что в идеальных системах информация передается безошибочно, а в реальных — с конечной вероятностью ошибки.

Отличия между достигнутыми и теоретически возможными показателями качества работы систем обусловлены рядом причин, некоторые из которых ранее уже упоминались. К числу таких причин в первую очередь относятся неидеальность характеристик реальных систем, приводящая к дополнительным искажениям сообщений, и неоптимальность используемых методов и способов передачи и приема информации.

1.8. КРАТКИЕ ИТОГИ ГЛ. 1

1. Непрерывно возрастающие потоки разнообразной информации и необходимость автоматизированной ее обработки привели к быстрому развитию цифровых систем связи. Такие системы обладают высокими качественными показателями по помехоустойчивости и скорости передачи, обеспечивают возможности создания сложных автоматизированных систем и комплексов.

2. Наиболее важными областями применения систем передачи цифровой информации являются передача данных, телеметрия, космическая связь и управление. В этих областях успешное решение многих задач без применения цифровых систем связи невозможно.

3. Качество передачи цифровой информации принято оценивать вероятностью ошибочного приема элемента цифрового сообщения и скоростью передачи. Для характеристики качества работы цифровой системы связи эти

показатели необходимо дополнять показателями удельных затрат полосы и энергии.

4. Причинами ухудшения качества передачи информации являются помехи и неидеальность характеристик как самой системы связи, так и среды распространения сигналов. Одна из основных задач теории и техники передачи информации состоит в отыскании таких методов и способов построения систем, которые сводили бы влияние указанных причин к минимуму. Это задача оптимизации линий связи.

5. Оптимизация линий связи представляет собой очень сложную проблему, решение которой зависит не только от степени развития теории, но и от уровня технических возможностей реализации оптимальных систем.

В настоящее время наиболее развитым направлением является оптимизация приемной части системы связи.

6. Теория оптимальных методов приема открывает большие возможности перед разработчиками новых систем передачи информации и позволяет получить ответ на ряд важных вопросов. К их числу относятся выбор структуры приемной части системы, оценка качества передачи информации, пути совершенствования реальных систем связи и т. д.

7. Исследования показывают, что все системы передачи цифровой информации можно разделить на три группы: 1) системы с малыми затратами полосы; 2) системы с малыми затратами энергии; 3) системы с малыми обобщенными затратами полосы и энергии. В системах первой группы необходимо использовать простые многоосновные сигналы, в системах второй группы — сложные сигналы.

8. Показатели качества работы современных цифровых систем связи еще заметно уступают предельно возможным показателям оптимальных систем. Это свидетельствует о том, что имеются значительные резервы и возможности совершенствования реальных систем связи.

КОНТРОЛЬНЫЕ ВОПРОСЫ И ЗАДАЧИ

1.1. Назвать основные причины быстрого развития систем передачи цифровой информации? Какие обстоятельства способствуют этому развитию?

1.2. Охарактеризовать основные особенности цифровой передачи информации.

1.3. Показать, что для обеспечения двухсторонней связи между большим количеством объектов ($N \gg 1$) необходимо иметь число линий связи, равное N^2.

1.4. Объяснить разницу между технической и информационной скоростью передачи информации. В каких случаях техническая скорость может быть больше информационной, а в каких меньше?

1.5. Назвать основные области применения систем передачи цифровой информации и дать их характеристику.

1.6. Записать математическую модель процесса передачи цифровой информации, когда операции кодирования и декодирования отсутствуют.

1.7. Записать выражение для оператора идеальной среды в двух случаях: 1) расстояние между передатчиком и приемником не меняется (стационарная линия связи); 2) расстояние между передатчиком и приемником меняется (динамическая линия связи).

1.8. Какие задачи должна решать эффективная теория оптимизации систем связи? Какие пути решения этих задач известны в настоящее время?

1.9. В чем суть проблемы оптимизации систем связи в целом?

1.10. В чем суть проблемы оптимизации приемной части системы связи?

1.11. Показать, что минимум удельных затрат энергии в идеальном гауссовском канале связи определяется величиной $\beta_E = \ln 2 \approx 0{,}7$.

1.12. Почему при передаче информации нельзя добиться одновременного уменьшения удельных затрат полосы и энергии?

1.13. Назвать характерные особенности простых и сложных сигналов. В каких случаях целесообразно применять те или иные сигналы?

1.14. Проанализировать зависимость обобщенных удельных затрат от удельных затрат полосы для идеальных систем по Шеннону, если эта зависимость имеет вид

$$1)\ \gamma = \alpha\beta_{\Delta f} + (1-\alpha)\beta_E,\quad 2)\ \gamma = \sqrt{\alpha\beta^2_{\Delta f} + (1-\alpha)\beta^2_E},$$

где α — весовой коэффициент, принимающий значения в интервале $0 \leqslant \alpha \leqslant 1$.

СПИСОК ЛИТЕРАТУРЫ

1. **Передача** данных. М., «Связь», 1969. Авт.: Е. В. Базилевич, В. С. Гуров, Н. Н. Етрухин, А. Г. Панкратов.

2. **Гуров В. С., Емельянов Г. А., Етрухин Н. Н.** Передача дискретной информации и телеграфия. М., «Связь», 1969.

3. **Передача** цифровой информации. Пер. с англ. М., ИЛ, 1963.

4. **Каналы** передачи данных. Под ред. В. О. Шварцмана. М., «Связь», 1970.

5. **Современная** радиолокация. Пер. с англ., М., «Сов. радио», 1969.

6. **Ван дер Зил А.** Шум. Источники, описание, измерение. Пер. с англ. М., «Сов. радио», 1973.

7. **Крэсснер Г. Н., Михаелс Дж. В.** Введение в системы космической связи. Пер. с англ., М., «Связь», 1967.

8. **Сиверс А. П.** Основы космической радиоэлектроники. М., «Сов. радио», 1969.

9. **Цифровые методы** в космической связи. Под ред. С. Голомба. Пер с англ, М., «Связь», 1969.

10. **Мановцев А. П.** Введение в цифровую радиотелеметрию. М., «Энергия», 1967.

11. **Воздшно-космическая** телеметрия. Пер. с англ. М., Воениздат, 1968.

12. **Достижения** в области телеметрии. Пер. с англ. М., «Мир», 1970.

13. **Калинин А. И., Черенкова Е. Л.** Распространение радиоволн и работа радиолиний. М., «Связь», 1971.

14. **Радиолинии** ионосферного рассеяния метровых волн. Под ред. Н. Н. Шумской. М., «Связь», 1973.

15. **Давыденко Ю. И.** Дальняя тропосферная связь. М., Воениздат, 1968.

16. **Сандерс Р. В.** Сравнение эффективности некоторых систем связи. — «Зарубежная радиоэлектроника», 1960, № 12.

17. **Миддлтон Д.** Очерки теории связи. Пер. с англ. М., «Сов. радио», 1966.

18. **Шеннон К.** Работы по теории информации и кибернетике. Пер. с англ. М., ИЛ, 1963.

19. **Бородин Л. Ф.** Введение в теорию помехоустойчивого кодирования. М., «Сов. радио», 1968.

20. **Шастова Г. А., Коёкин А. И.** Выбор и оптимизация структуры информационных систем. М., «Энергия», 1972.

21. **Котельников В. А.** Теория потенциальной помехоустойчивости. М., Госэнергоиздат, 1956.

22. **Шеннон К., Галагер Р., Берлекамп Е.** Нижние границы вероятности ошибки кодирования в дискретных каналах без памяти. — «Зарубежная радиоэлектроника», 1968, № 2, 6.

23. **Возенкрафт Дж., Джекобс И.** Теоретические основы техники связи. М., «Мир», 1969.

24. **Харкевич А. А.** Спектры и анализ. М., Изд-во технико-теоретической литературы, 1957.

25. **Сервинский Е. Г.** Оптимизация систем передачи дискретной информации. М., «Связь», 1974.

26. **Гуткин Л. С.** Оптимизация радиоэлектронных устройств. М., «Сов. радио», 1975.

Глава 2

Представление информации в цифровой форме

2.1. ВВЕДЕНИЕ

Дискретными называют сообщения, образованные из отдельных элементов (символов, букв, цифр, импульсов и т. п.), принимающих конечное число различных значений. Как уже говорилось ранее, эти элементы образуют алфавит дискретного источника. Любой источник, располагающий алфавитом из конечного числа элементов, может образовать только конечное число различных дискретных сообщений заданной длительности.

Таким образом, дискретное сообщение представляет собой последовательность элементов, каждый из которых может принимать лишь конечное число различных значений. Дискретное сообщение называют цифровым, если каждому его элементу (или некоторой комбинации этих элементов) приписывают определенное цифровое значение. Примером дискретных сообщений являются телеграфные сообщения, цифровые данные, буквенные тексты, разовые команды в некоторых системах телеуправления и т. п. Дискретность подобных сообщений обусловлена самой природой источников, вырабатывающих эти сообщения.

Однако иногда дискретные сообщения являются результатом преобразования непрерывной (аналоговой) информации в цифровую форму. Как уже указывалось в гл. 1, в настоящее время происходит непрерывное расширение областей применения систем передачи цифровой информации и все большее число различных видов аналоговой информации стремятся передавать в цифровой форме. Это относится к передаче телефонных сообщений, фотоизображений, данных телеметрии и т. п. Таким образом, дискретные сообщения могут быть как первичными, так и вторичными, полученными из непрерывных.

Независимо от природы возникновения все дискретные сообщения, подлежащие передаче по линии связи, должны быть преобразованы в цифровую форму, наи-

более удобную для передачи и приема. При этом стремятся как можно полнее учесть требования стандартизации и унификации параметров и характеристик используемых линий и систем связи. Особенно это касается линий и систем передачи данных.

К качеству передачи разных дискретных сообщений предъявляются различные требования. Так, при передаче команд, как правило, предъявляются очень высокие требования к достоверности и сравнительно невысокие требования к скорости передачи (вероятность искажения команды иногда не должна превышать 10^{-6}—10^{-8}, а время передачи команды может быть значительным — порядка секунды). Дискретные команды в некоторых системах управления должны обладать высокой криптостойкостью (не поддаваться быстрой расшифровке) и имитостойкостью (обладать свойствами, затрудняющими повторение этих команд с целью нарушения нормальной работы системы управления). Как уже было выяснено в гл. 1, для удовлетворения указанных требований подходят сложные сигналы с большой базой.

При передаче цифровых последовательностей, полученных дискретизацией непрерывных сообщений, требования к достоверности передачи ниже, чем при передаче команд, и допустимая вероятность ошибки кодовой комбинации обычно составляет 10^{-3}—10^{-5}. Зато требования к скорости передачи существенно выше, чем при передаче команд, и могут составлять десятки и сотни тысяч двоичных единиц в секунду.

Как уже неоднократно указывалось, для передачи непрерывных сообщений по цифровым линиям связи необходимо преобразовать эти сообщения в цифровую форму. Такое преобразование связано с выполнением трех операций: 1) дискретизации сообщений по времени; 2) дискретизации сообщений по уровню (квантование); 3) превращения сообщений, дискретизированных по времени и по уровню, в последовательность чисел, выраженных в виде соответствующих кодовых комбинаций.

В принципе не имеет значения, в каком порядке следует выполнять первые две операции, однако технически удобнее сначала осуществить дискретизацию сообщений по времени, а затем уже квантование. Обычно

квантование выполняется одновременно с получением кодовых комбинаций.

Устройства, которые осуществляют преобразование непрерывных сообщений в цифровую форму, называют преобразователями «аналог—цифра». Их часто называют также преобразователями «напряжение—код».

Учитывая принципиальную важность теории преобразования непрерывных сообщений в цифровую форму, рассмотрим некоторые вопросы этой теории более детально.

2.2. ДИСКРЕТИЗАЦИЯ НЕПРЕРЫВНЫХ СООБЩЕНИЙ ПО ВРЕМЕНИ

2.2.1. Постановка задачи. Принципы дискретизации по времени

Суть дискретизации по времени состоит в том, что непрерывное сообщение заменяется последовательностью его мгновенных значений (отсчетов), взятых в дискретных точках времени (рис. 2.1). При такой замене из рассмотрения исключается все множество значений непрерывной функции времени, находящихся внутри интервалов времени Δt. Полученная при этом функция имеет вид последовательности отсчетов, взятых в дискретные моменты времени. Часто такую функцию называют решетчатой.

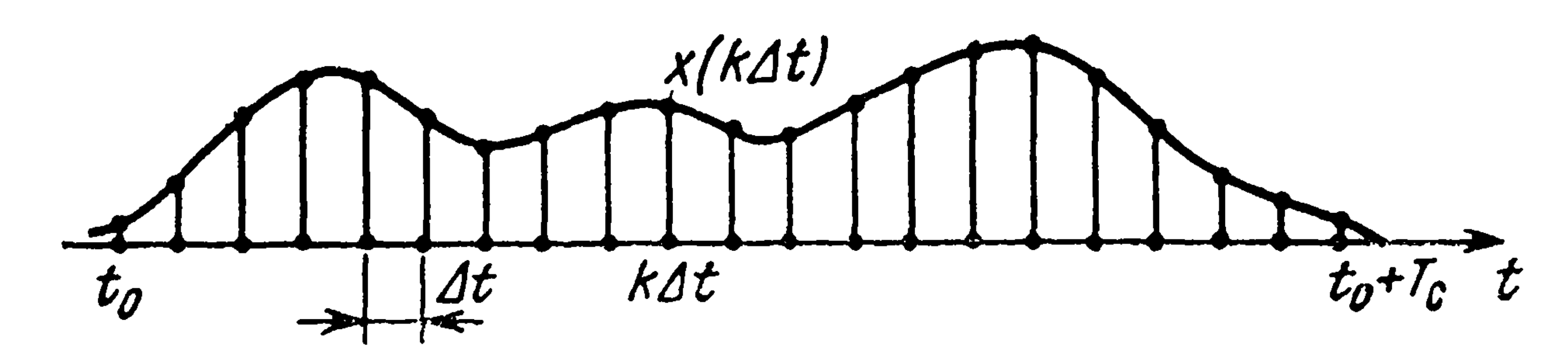


Рис. 2.1.

Дискретизация по времени может быть равномерной (принудительной), когда интервал дискретизации Δt остается неизменным, и неравномерной, когда интервал Δt меняется в соответствии с изменением тех или иных характеристик сообщения, например, в соответствии со скоростью изменения сообщения во времени (адаптивная, или приспосабливающаяся дискретизация). Теоретическое рассмотрение и практическая реализация адаптивной дискретизации представляют значительные трудности, так как для такой дискретизации необходимо располагать большим объемом априорных сведений о дискретизируемых сообщениях и существенно усложнять устройства дискретизации. Поэтому такая дискретизация пока не получила широкого распространения.

В настоящее время наиболее широко применяется равномерная дискретизация. Теория и техника такой дискретизации достаточно

хорошо разработаны, она сравнительно проста и удобна для решения многих задач. В частности, равномерная дискретизация может использоваться (и используется) как промежуточная операция при решении задач адаптивной дискретизации, сжатия сообщений для устранения в них изобыточности, временного уплотнения каналов в системах связи и т. п. Поэтому ниже будут рассмотрены основные особенности только равномерной дискретизации. С некоторыми вопросами адаптивной дискретизации и ее приложениями можно ознакомиться в работах [1—3], которые содержат также большую библиографию по этим вопросам.

В основе математического описания дискретизации непрерывных функций по времени лежит так называемая импульсная функция дискретизации $a_д(t)$, которая представляет собой периодическую последовательность δ-функций (единичных импульсных функций), следующих через интервалы времени Δt *:

$$a_д(t) = \sum_{k=-\infty}^{\infty} \delta(t - k\Delta t). \tag{2.1}$$

Такую функцию можно назвать также периодической стробирующей или коммутационной импульсной функцией. В работе [4] подобная функция названа выборочной.

Используя преобразование Фурье, можно показать, что спектр импульсной функции дискретизации определяется выражением

$$A_д(f) = \frac{1}{\Delta t} \sum_{k=-\infty}^{\infty} \delta\left(f - \frac{k}{\Delta t}\right), \tag{2.2}$$

которое представляет собой периодическую последовательность δ-функций, следующих через частотные интервалы $\Delta f = 1/\Delta t$.

Заметим, что в соответствии со свойством δ-функций

$$\int_{-\infty}^{\infty} \delta(\xi - \xi_0) d\xi = 1,$$

площадь каждой составляющей в выражении (2.1) равна единице, а в выражении (2.2) — величине $1/\Delta t$. Эти значения площадей определяют «веса» δ-функций.

Дискретизация непрерывной функции времени $x(t)$ с математической точки зрения представляет собой умножение этой функции на функцию $a_д(t)$:

$$x_д(t) = x(t) a_д(t) = \sum_{k=-\infty}^{\infty} x(t) \delta(t - k\Delta t). \tag{2.3}$$

В соответствии с фильтрующим свойством δ-функции, которое можно выразить соотношением

$$\int_{-\infty}^{\infty} x(\xi) \delta(\xi - \xi_0) d\xi = x(\xi_0), \tag{2.4}$$

---

* С теорией и свойствами δ-функций можно ознакомиться в работах [4, 5].

имеем

$$x(t)\delta(t-k\Delta t)=x(k\Delta t)\delta(t-k\Delta t). \qquad (2.5)$$

Это выражение означает, что умножение функции $x(t)$ на единичный δ-импульс приводит к тому, что площадь этого импульса становится равной вместо единицы значению функции $x(t)$ в момент времени $t=k\Delta t$. Эту площадь обычно называют весовым коэффициентом (весом) δ-импульса. Иными словами, умножение $x(t)$ на единичный δ-импульс соответствует получению отсчета функции $x(t)$ в момент $t=k\Delta t$.

С учетом (2.5) выражение (2.3) принимает вид

$$x_{д}(t)=\sum_{k=-\infty}^{\infty} x(k\Delta t)\delta(t-k\Delta t). \qquad (2.6)$$

Следовательно, умножение сообщения $x(t)$ на импульсную функцию дискретизации приводит к образованию периодической последовательности δ-импульсов, веса которых равны мгновенным значениям сообщения в моменты времени $t=k\Delta t$, т. е. в моменты взятия отсчетов.

На практике реализовать импульсную функцию дискретизации в виде δ-импульсов невозможно. Технически дискретизация непрерывного сообщения реализуется ключевыми устройствами, управляемыми периодической последовательностью коротких прямоугольных импульсов. При этом длительность отсчетов конечна, поскольку отсчет берется не в одной точке, а в некотором интервале времени, равном длительности импульса τ. Если выбрать величину τ так, чтобы обеспечивалось условие $\tau/\Delta t \ll 1$, образуется последовательность коротких импульсов, амплитуды которых пропорциональны мгновенным значениям сообщения.

Практическую реализацию дискретизации по времени часто называют импульсным преобразованием непрерывного сообщения. По сути дела такое преобразование эквивалентно получению амплитудно-модулированной последовательности импульсов (АИМ). При этом, если среднее значение непрерывного сообщения $x(t)=0$, получается так называемая балансная АИМ (биполярная последовательность), если же среднее значение $x(t)$ равно половине полной шкалы изменения величины x, получается модулированная последовательность импульсов, соответствующая АИМ с глубиной модуляции $m_a=1$ (униполярная последовательность).

При решении задач дискретизации непрерывных сообщении по времени, возникает ряд вопросов: 1) из каких соображений необходимо исходить при выборе интервала дискретизации Δt; 2) какова точность замены непрерывного сообщения последовательностью его отсчетов, взятых в дискретные моменты времени, и от чего она зависит; 3) каков максимально допустимый интервал дискретизации Δt, при котором еще принципиально возможно восстановление непрерывного сообщения по его отсчетам.

Получить ответ на эти и другие вопросы можно лишь в случае, если проблему дискретизации по времени рассматривать в неразрывной связи с обратной проблемой — восстановлением непрерывной функции времени по ее мгновенным значениям, известным только в дискретные моменты времени. Сложность этой проблемы состоит в том, что нужно восстановить утраченные при дискретизации

сведения о поведении непрерывной функции времени в промежутках между отсчетами. Очевидно, что чем меньшим количеством отсчетов заменяется сообщение длительностью T_c, тем продолжительнее интервал дискретизации Δt и тем сложнее выполнить восстановление исходной функции. И наоборот, чем больше отсчетов, тем короче интервал) Δt и тем проще оказывается восстановление *.

Таким образом, рассматриваемая проблема дискретизации не имеет однозначного решения. Осуществление более экономного импульсного преобразования непрерывного сообщения связано с усложнением задачи восстановления этого сообщения по его отсчетам. И наоборот, неэкономное (избыточное) импульсное преобразование приводит к упрощению процедуры восстановления.

При реализации импульсного преобразования приходится принимать компромиссное решение, учитывающее отмеченные выше особенности.

Выбор интервала дискретизации Δt является одним из принципиально важных вопросов теории и техники передачи сообщений.

2.2.2. Математическая модель непрерывного сообщения

Сложность задачи выбора интервала дискретизации состоит в том, что при ее решении необходимо учитывать свойства исходных непрерывных сообщений, способ восстановления этих сообщений по отсчетам и требуемую точность восстановления.

Первым этапом решения является выбор модели непрерывного сообщения, представляющей собой некоторую математическую идеализацию такого сообщения. От удачного выбора модели зависит простота теории и возможность ее приложений к достаточно широкому кругу вопросов. Чтобы не усложнять задачу, в модели необходимо сохранить лишь те характерные особенности сообщения, которые имеют наиболее существенное значение для рассматриваемых задач.

Непрерывные реальные сообщения с математической точки зрения необходимо рассматривать как реализации некоторого нестационарного случайного процесса Однако теория таких процессов сложна, разработана еще недостаточно и, несмотря на большую важность, не нашла пока широкого применения в инженерной практике.

Во многих практически интересных и важных случаях нестационарность реальных сообщений проявляется не очень резко (например, телеметрическая информация в стационарных режимах, речевые сообщения и т. п.). Статистические характеристики таких процессов (средние значения, дисперсия и т п) изменяются заметно только на коротких интервалах времени, а на значительной части интервала существования рассматриваемого процесса практически не меняются. Благодаря такому положению можно

* Замена непрерывного сообщения длительностью T_c конечным числом отсчетов принципиально не может быть точной. При такой замене всегда появляется погрешность Поэтому правильнее говорить о замене с некоторой заранее установленной погрешностью Этот вопрос обсуждается ниже.

ввести идеализацию многих реальных процессов в виде стационарных или кусочно-стационарных процессов.

Удобство такой идеализации заключается в том, что при рассмотрении многих вопросов теории импульсного преобразования можно пользоваться хорошо разработанной теорией стационарных случайных процессов. Кроме того, многие стационарные процессы обладают ценным свойством эргодичности, которое позволяет получить характеристики процесса по одной его реализации [5]. Любые реальные процессы имеют конечную протяженность во времени, поэтому свойство эргодичности для таких процессов нарушается и определение характеристик процесса по одной реализации становится неточным. Однако при условии, что

$$T_c \Delta F_э \gg 1, \tag{2.7}$$

где T_c — длительность процесса, а $\Delta F_э$ — ширина его энергетического спектра, указанная неточность невелика и с нею в ряде случаев можно не считаться.

Условие (2.7) является одним из необходимых условий квазистационарности случайного процесса. Физически оно означает, что непрерывное сообщение за время T_c должно много раз менять знак производной *.

Для характеристики стационарного случайного процесса $\{x(t)\}$, выбранного в качестве модели непрерывных сообщений, необходимо знать его многомерную функцию плотности распределения. Эта функция часто неизвестна, и процесс обычно характеризуют энергетическим спектром $G_x(\Omega)$ или корреляционной функцией $\varphi_x(\tau)$.

Как известно, энергетическим спектром случайного процесса называют функцию, которая определяет спектральную плотность мощности этого процесса. Заметим, что при проведении теоретических исследований энергетический спектр обычно определяют во всей области значений частот (как положительных, так и отрицательных). Так как $G_x(\Omega)$ четная функция:

$$G_x(-\Omega) = G_x(\Omega),$$

то в тех случаях, когда это удобно, можно пользоваться понятием одностороннего энергетического спектра, который равен $2G_x(\Omega)$ для всех положительных значений частоты.

Функция $G_x(\Omega)$ может быть выражена через спектр временного процесса. Если амплитудный спектр одной из реализаций процесса $x(t)$ на интервале T_c равен $A_x(\Omega)$, можно показать, что

$$G_x(\Omega) = \lim_{T_c \to \infty} (\overline{|A_x(\Omega)|^2}/T_c), \tag{2.8}$$

где $|A_x(\Omega)|$ — модуль амплитудного спектра; черта сверху означает усреднение по множеству реализаций (статистическое усреднение).

Для квазистационарных процессов конечной длительности выражение (2.8) становится приближенным, однако при выполнении условия (2.7) с этим можно не считаться.

* Если реальный процесс заменен кусочно-стационарным, то условие (2.7) должно выполняться для каждого частичного интервала квазистационарности T_i, входящего в состав общего интервала T_c.

На процесс могут быть наложены некоторые ограничения, учитывающие физические свойства реальных сообщений, в частности:

а) конечное значение средней мощности процесса

$$P_x = \frac{1}{T_c}\int\limits_0^{T_c} x^2(t)dt < \infty, \qquad (2.9)$$

для стационарного эргодического процесса эта величина равна

$$P_x = \frac{1}{\pi}\int\limits_0^{\infty} G_x(\Omega)d\Omega;$$

б) конечная шкала мгновенных значений

$$\max |x(t)| \leqslant X_m. \qquad (2.10)$$

Помимо указанных идеализаций и ограничений, при выборе модели непрерывных сообщений часто вводят еще одну принципиально важную идеализацию, связанную с ограничением энергетического спектра по частоте. Суть этой идеализации состоит в следующем. При передаче непрерывных сообщений по каналам связи всегда имеются неизбежные помехи. Так как спектральная плотность энергетического спектра любого реального сообщения уменьшается с ростом частоты, то, начиная с некоторой частоты, спектральная плотность сообщения оказыватся меньше спектральной плотности помех и, следовательно, такие составляющие спектра сообщения почти не вносят вклада в полезную информацию. Это обстоятельство позволяет ввести идеализацию, заключающуюся в том, что непрерывные сообщения рассматриваются как случайный процесс с ограниченным энергетическим спектром.

Среднеквадратичная ошибка, связанная с ограничением энергетического спектра стационарного случайного процесса граничной частотой $\Omega_в$, определяется выражением

$$\delta^2_F = \frac{\int\limits_{\Omega_в}^{\infty} G_x(\Omega)d\Omega}{\int\limits_0^{\infty} G_x(\Omega)\ \Omega} = \frac{\Delta P_x}{P_x} = \frac{\Delta E_x}{E_x}, \qquad (2.11)$$

которое можно трактовать как отношение средней мощности ΔP_x (или энергии ΔE_x) отброшенной части спектра к средней мощности P_x (или энергии E_x) всего спектра.

Итак, в качестве модели непрерывных сообщении обычно выбирают случайный квазистационарный процесс $\{x(t)\}$, имеющий конечную протяженность во времени и ограниченный энергетический спектр, а также удовлетворяющий условиям (2.9) и (2.10).

Из теории преобразования Фурье известно, что условия ограниченности функции по времени существования и по спектру одновременно выполняться не могут Это означает, что принятая модель для сообщений конечной длительности некорректна с математиче-

ской точки зрения. Многочисленные исследования [6, 7] показали, что случайный процесс с ограниченным спектром не является точной моделью непрерывных сообщений и с информационной точки зрения. Вместе с тем, исследования показали, что при соответствующем выборе частоты ограничения спектра переход от реальных сообщений к их идеализированной модели не связан с большими погрешностями и во многих случаях вполне допустим.

Рассмотренная модель проста, наглядна и удобна для решения ряда задач, возникающих при дискретизации непрерывных сообщений по времени. Поэтому, несмотря на отмеченные выше принципиальные недостатки, такая модель получила очень широкое распространение.

2.2.3. Предельная дискретизация по времени непрерывных сообщений с ограниченным спектром

В теории дискретизации непрерывных сообщений особую важность приобретает вопрос о максимальном (предельном) интервале дискретизации $\Delta t = \max \Delta t$, при котором еще принципиально возможно восстановить непрерывную функцию времени с заданной точностью по ее отсчетам. Необходимое число отсчетов при этом будет минимальным и определится величиной

$$\min m = [T_c / \max \Delta t],$$

где скобки указывают на то, что берется целая часть отношения.

Дискретизацию непрерывных сообщений, соответствующую такому условию, назовем предельной. Она обеспечивает представление непрерывного сообщения с заданной точностью минимальным (безызбыточным) количеством отсчетов. В некоторых работах такую дискретизацию называют оптимальной [2] *.

Проблема предельной дискретизации сложна и, несмотря на значительное количество исследований, далека от завершения.

Большое внимание, которое уделяется теории предельной дискретизации по времени, объясняется тем, что она лежит в основе анализа ряда важнейших вопросов преобразования и передачи непрерывных сообщений, в частности, при исследовании вопросов устранения избыточности сообщений («сжатие информации»), пропускной способности каналов и т. п.

В настоящее время наиболее разработанной и широко применяемой предельной дискретизацией непрерывных сообщений является дискретизация, основанная на теореме Котельникова.

Остановимся на исходных положениях этой теоремы. Как уже указывалось выше, при построении теории дискретизации необходимо опираться на некоторую модель сообщения. Допустим, что нам известна одна из реализаций $x(t)$ квазистационарного случай-

* Такое определение в данном случае нельзя считать удачным, поскольку выбор числа отсчетов в качестве критерия оптимальности не учитывает проблемы интерполяции. Оптимизация дискретизации должна выполняться по таким критериям, которые бы учитывали обе процедуры — дискретизацию и интерполяцию.

ного процесса $\{x(t)\}$, соответствующего совокупности возможных непрерывных сообщений. Как бы сложно эта реализация ни выглядела, она представляет собой некоторую неслучайную (детерминированную) функцию времени. Пользуясь преобразованием Фурье, можно найти комплексный амплитудный спектр этой функции:

$$x(t) \Longleftrightarrow A_x(j\Omega) = A_x(\Omega)e^{j\varphi_\Omega} = \int_0^{T_c} x(t)e^{-j\Omega t}dt. \qquad (2.12)$$

Если функция $x(t)$ задана графически в виде фотограммы или существует в виде записи на магнитную ленту, можно воспользоваться одним из известных методов вычислительной математики и построить график амплитудного спектра. Исходя из требуемой точности и допустимого уровня помех, ограничим этот спектр некоторой частотой $\Omega_в = 2\pi F_в$, т. е. введем модель сообщения с ограниченным амплитудным спектром. Введенная модель характеризуется следующими особенностями: 1) относится к одной реализации случайного процесса, т. е. соответствует детерминированной функции; 2) имеет ограниченный спектр.

Для такой модели верна следующая теорема: если непрерывная функция времени $x(t)$ имеет спектр, ограниченный полосой частот от нуля до $F_в$, то эта функция полностью определяется последовательностью своих мгновенных значений, взятых в моменты времени, отсчитываемые через интервалы $\Delta t = 1/2F_в$.

Эта теорема как математическое положение установлена давно (см. например, [8]). Однако в наиболее четкой форме ее сформулировал, доказал и применил к конкретным проблемам передачи сообщений В. А. Котельников в 1933 г. Поэтому вполне справедливо в отечественной литературе указанную теорему называют теоремой Котельникова *. В иностранной литературе (в основном американской и английской) подобную теорему обычно называют теоремой отсчетов или теоремой Найквиста **.

Доказательство теоремы Котельникова содержится в ряде книг [8, 9, 12], и мы им заниматься не будем. Результатом доказательства теоремы является выражение

$$x(t) = \sum_{k=-\infty}^{\infty} x(k\Delta t)\,\frac{\sin \Omega_в(t - k\Delta t)}{\Omega_в(t - k\Delta t)}, \qquad (2.13)$$

где

$$\Delta t = 1/2F_в = \pi/\Omega_в. \qquad (2.14)$$

Разложение непрерывной функции времени $x(t)$ в ряд вида (2.13) очень важно и приобрело в теории информации и других областях не меньшее значение, чем разложение Фурье. В этом раз-

* Формулировка теоремы, данная ее автором, несколько отличается от приведенной выше, но суть ее та же (см. [9]).

** В 1928 г. теорема была сформулирована без доказательства английским ученым Найквистом [10]. В 1948 г. американский ученый Шеннон при разработке математической теории связи вновь ввел и доказал эту теорему, назвав ее теоремой отсчетов [11].

ложении значения $x(k\Delta t)$ в дискретных точках времени можно рассматривать как координаты x_k, а функции вида

$$\sin \Omega_в(t-k\Delta t)/\Omega_в(t-k\Delta t)$$

— как базисные функции $\varphi_k(t)$. Тогда выражение (2.13) можно рассматривать как частный случай обобщенного ряда Фурье:

$$x(t)=\sum_{k=-\infty}^{\infty} x_k\varphi_k(t). \quad (2.15)$$

Принципиальная важность выражения (2.13) заключается в том, что оно дает решение как прямой задачи (выбор интервала дискретизации Δt), так и обратной задачи (восстановление непрерывной функции $x(t)$ по значениям ее отсчетов). В соответствии с выражением (2.15) процедура восстановления исходной функции сводится к суммированию бесконечного числа функций $\varphi_k(t)$ с весовыми коэффициентами, равными отсчетам x_k. Это означает, что точное восстановление функции $x(t)$ с ограниченным спектром возможно только при бесконечной протяженности этой функции во времени. В действительности же все реальные сообщения имеют ограниченную продолжительность во времени. Характерной особенностью реальных сообщений является то, что они относятся к такому классу функций, у которых почти вся энергия сосредоточена в конечных интервалах времени и полосы частот.

Если у сообщения длительностью $T_с$ ограничить спектр на частоте $F_в$, то в соответствии с теоремой Котельникова можно образовать число отсчетов, равное *

$$m=T_с/\Delta t=2F_вT_с. \quad (2.16)$$

В этом случае ряд (2.15) будет содержать конечное число членов и, следовательно, представление непрерывной функции таким рядом будет неточным:

$$x(t)\approx\hat{x}(t)=\sum_{k=-m/2}^{m/2} x_k\varphi_k(t). \quad (2.17)$$

Приближенность этого выражения проявляется в том, что при конечном числе членов ряда их сумма точно совпадает с мгновенными значениями функции $x(t)$ не на всем интервале времени $T_с$, а только в точках отсчетов. В интервалах между точками отсчета значения функции $x(t)$ и функции приближения $\hat{x}(t)$ различаются и появляется погрешность. Уменьшить эту погрешность можно путем увеличения числа членов ряда. Как следует из выражения (2.16), при конечной длительности сообщения $T_с$ это можно сделать, только уменьшая интервал дискретизации Δt, т. е. увеличивая значение частоты $F_в$, которой ограничивается спектр сообщения.

---

* Строго говоря, $m=T_с/\Delta t+1=2F_вT_с+1$, но так как обычно $m\gg1$, то единицей пренебрегают. Естественно, что число отсчетов всегда целое, поэтому при определении m берется только целая часть полученного значения.

Среднеквадратичную ошибку, вызванную указанной погрешностью, можно определить соотношением

$$\delta^2_T = \frac{\int\limits_0^{T_c} [x(t) - \widehat{x}(t)]^2 dt}{\int\limits_0^{T_c} x^2(t) dt} = \frac{\int\limits_0^{T_c} \varepsilon^2_T(t) dt}{E_x}, \qquad (2.18)$$

где $\varepsilon_T(t)$ — погрешность между $x(t)$ и $\hat{x}(t)$; E_x — энергия непрерывного сообщения.

Основываясь на известном в теории разложений функции положении о том, что при одинаковом числе членов разложения наименьшую среднеквадратичную ошибку обеспечивает ряд Фурье [9], можно показать, что ошибка, определяемая выражением (2.18), не может быть меньше ошибки, определяемой выражением (2.11).

Таким образом, дискретизация непрерывного сообщения конечной длительности в соответствии с теоремой Котельникова связана с ошибкой, одна составляющая которой обусловлена введением модели с ограниченным спектром, а вторая — учетом конечного числа членов ряда разложения. Общая среднеквадратичная ошибка при дискретизации непрерывного сообщения конечной длительности может быть определена выражением

$$\delta^2_д = \delta^2_T + \delta^2_F \geqslant 2\delta_F = \frac{2}{E_x} \int\limits_{\Omega_в}^{\infty} G(\Omega) d\Omega = 2\Delta E_x / E_x. \qquad (2.19)$$

Заметим, что в работах [6, 8] даются оценки ошибки дискретизации по времени, близкие к записанной выше. В частности, в работе [6] приводятся нижняя и верхняя оценки в виде

$$\Delta E_x / E_x < \delta^2_д < 3\Delta E_x / E_x. \qquad (2.20)$$

Восстановление непрерывной функции времени с конечной длительностью по ее отсчетам должно выполняться в соответствии с выражением (2.13). Эта процедура может быть выполнена двумя способами: фильтрационным с применением аналогового фильтра и интерполяционным с применением специальных интерполяторов или универсальных вычислительных машин.

Рассмотрим кратко суть этих способов.

При фильтрационном способе восстановления последовательность отсчетов, определяемая выражением (2.13), в котором интервал дискретизации в соответствии с теоремой Котельникова равен $\Delta t = 1/2F_в$, подается на фильтр нижних частот. Напряжение на выходе фильтра определяется суперпозицией откликов фильтра на каждый поступающий отсчет. Нетрудно показать, что для получения выходного напряжения в виде, определяемом выражением (2.13), необходимо применить фильтр нижних частот, импульсная переходная функция которого должна иметь вид

$$h(t) = 2F_в \sin \Omega_в t / \Omega_в t.$$

Этой функции соответствует фильтр нижних частот с прямоугольной передаточной функцией. Как известно, такой фильтр физически нереализуем [4, 13].

При интерполяционном способе процедура восстановления должна выполняться в соответствии с выражением (2.13), которое можно рассматривать как алгоритм этой процедуры. В соответствии с этим алгоритмом необходимо создать $m=2F_{в}T_{c}$ функций вида

$$\varphi_k(t) = \sin \Omega_в(t - k\Delta t)/\Omega_в(t - k\Delta t)$$

и просуммировать их с учетом весовых коэффициентов, равных переданным отсчетам $x_k=x(k\Delta t)$. Очевидно, что технически это сложная операция, так как она требует запоминания m отсчетов, генерирования функций $\varphi_k(t)$ и их суммирования с учетом весовых коэффициентов. Подобный способ восстановления также не может быть физически реализован точно *.

По указанным причинам предельная дискретизация непрерывных сообщений в соответствии с теоремой Котельникова, а также последующее восстановление сообщений рассмотренными способами не находят практического применения. Основная ценность предельной дискретизации по времени состоит в том, что она определяет принципиальные возможности дискретизации и указывает пути, позволяющие приблизиться к этим возможностям.

Теорема Котельникова получила широкое применение в различных теоретических исследованиях. Однако необходимо отметить, что с точки зрения статистической теории связи эта теорема имеет и слабые стороны: она верна лишь для детерминированных функций с ограниченным спектром. Применение теоремы к случайным функциям времени требует специального рассмотрения. Эти вопросы многократно рассматривались в ряде работ. В частности, в [14] показано, что теорема строго верна лишь для случайных процессов с равномерным и ограниченным энергетическим спектром. Это утверждение становится понятным, если рассмотреть выражение (2.8). Действительно, энергетический спектр имеет такую же форму, как и амплитудный, только в одном случае — когда амплитудный спектр прямоуголен. Если амплитудный спектр неравномерный, то энергетический спектр становится еще более неравномерным. Для таких процессов дискретизация по Котельникову уже не предельна. Поэтому применение предельной дискретизации в соответствии с теоремой Котельникова можно считать приемлемым лишь для случайных процессов со спектрами, достаточно близкими к равномерным в пределах рассматриваемой области частот и быстро спадающими вне этой области.

2.2.4. Дискретизация по времени непрерывных сообщений с учетом реальных способов восстановления

Выше указывалось, что практическое применение теоремы Котельникова наталкивается на серьезные трудности. Это связано с тем, что в реальных устройствах не удается точно выполнить ряд

* Заметим, что при реализации рассмотренных способов, помимо неизбежной неточности, результат восстановления получается еще и с задержкой во времени, равной длительности исходного непрерывного сообщения.

важных условий, при которых эта теорема справедлива: 1) ограниченность спектра сообщения; 2) идеальность коммутирующей последовательности δ-функций. Эти условия однозначно определяют физически нереализуемый алгоритм восстановления исходного сообщения. Поэтому в соответствии с теоремой Котельникова практически невозможно реализовать точно ни дискретизацию непрерывного сообщения, ни его восстановление. Приближенная же реализация этих операций (с учетом требуемой точности восстановления сообщения) приводит к сложным схемным решениям алгоритма восстановления или к большому объему вычислений при использовании ЭВМ.

В настоящее время в технических приложениях выбор интервала дискретизации непрерывных сообщений определяется в значительной мере требованиями удобства и простоты восстановления этих сообщений по их отсчетам. На практике широкое распространение получил метод восстановления с помощью интерполяционного многочлена Лагранжа [2, 3, 8, 14]. Наиболее часто используется многочлен не очень высокого порядка (обычно не более второго). Это объясняется широким развитием методов автоматической обработки информации, основанных, как правило, а простых способах интерполяции. Наибольшее применение в настоящее время находит интерполяция нулевого, первого и второго порядков (ступенчатая, линейная и параболическая).

Естественно, что применение простых способов интерполяции требует выбора более высокой частоты дискретизации, чем это следует из теоремы Котельникова. Получающаяся в результате такой дискретизации последовательность отсчетов с информационной точки зрения обладает определенной избыточностью. Оценим величину этой избыточности.

Примем за меру избыточности отсчетов величину

$$r=(m_ф-m_0)/m_ф=1-m_0/m_ф,\quad 0\leqslant r\leqslant 1, \tag{2.21}$$

где m_0 — минимально возможное, а $m_ф$ — фактическое количество отсчетов, сделанных за время, определяемое длительностью сообщения T_c. Нетрудно показать, что принятая мера находится в полном соответствии с определением избыточности информации, широко применяемом в теории информации (см. например, [12]).

Частота дискретизации (частота повторения отсчетов) связана с интервалом дискретизации Δt и верхней граничной частотой спектра сообщения $F_в$ соотношением

$$F_п=1/\Delta t=\mu F_в, \tag{2.22}$$

где μ — коэффициент следования отсчетов.

Для предельной дискретизации по Котельникову имеем

$$\Delta t=\max\Delta t=1/2F_в;\quad m_0=2F_вT_c;\quad \mu=2. \tag{2.23}$$

Для реальной (фактически используемой) дискретизации $\mu>2$, и следовательно,

$$\Delta t=\Delta t_ф=1/\mu F_в;\ m_ф=\mu F_вT_c>m_0. \tag{2.24}$$

В технических приложениях избыточность отсчетов часто оценивают величиной

$$k_{изб}=\max\Delta t/\Delta t_ф=m_ф/m_0=\mu/2\geqslant 1, \tag{2.25}$$

которую называют коэффициентом избыточности.

С учетом соотношения (2.25) выражение (2.21) можно записать в виде

$$r = 1 - 2/\mu = 1 - 1/k_{\text{изб}}. \qquad (2.26)$$

Из выражения (2.25) или (2.26) следует, что, если большая избыточность нежелательна, необходимо выбирать величину μ близкой к предельному значению ($\mu \sim 2$), но при этом появляются указанные выше технические трудности. Стремление уменьшить эти трудности приводит к необходимости выбирать величину μ больше предельной ($\mu > 2$), но при этом увеличивается избыточность отсчетов. Следовательно, с инженерной точки зрения проблема импульсного преобразования непрерывных сообщений не имеет однозначного решения. Поэтому при ее технической реализации приходится принимать компромиссное решение. Желательно, чтобы при достаточно простых способах восстановления избыточность отсчетов не была слишком большой.

Основываясь на результатах теории интерполяции, в работе [15] показано, что для сообщений с прямоугольным (или близким к нему) спектром интервал дискретизации равен: при ступенчатой интерполяции

$$\Delta t_{\text{ст}} = \frac{\sqrt{3}\delta_0}{\Omega_{\text{в}}} = \frac{\sqrt{3}}{\pi}\delta_0 \Delta t_{\text{пр}} \approx 0{,}55\delta_0 \Delta t_{\text{пр}}; \qquad (2.27)$$

при линейной интерполяции

$$\Delta t_{\text{л}} = 2\sqrt{2\sqrt{5}\delta_0}/\Omega_{\text{в}} \approx 1{,}35\sqrt{\delta_0}\Delta t_{\text{пр}}; \qquad (2.28)$$

при параболической интерполяции

$$\Delta t_{\text{пб}} = \sqrt[3]{15{,}5\sqrt{7}\delta_0}/\Omega_{\text{в}} \approx \sqrt[3]{\delta_0}\Delta t_{\text{пр}} \qquad (2.29)$$

В этих формулах $F_{\text{в}}$ — верхняя граничная частота спектра сообщения; $\Delta t_{\text{пр}} = 1/2F_{\text{в}}$ — предельный интервал дискретизации; δ_0 — допустимая относительная погрешность равномерного приближения, определяемая выражением

$$\delta_0 = \max|x(t) - \hat{x}(t)|/X_m = \Delta x_0/x_{\max}. \qquad (2.30)$$

В свою очередь, здесь $x(t)$ — исходное непрерывное сообщение; $\hat{x}(t)$ — сообщение, восстановленное по отсчетам с помощью рассматриваемого вида интерполяции; $X_m = x_{\max}$ — максимальное значение интервала изменения исходного сообщения (полная шкала изменения мгновенных значений процесса).

Приведенные результаты показывают, что избыточность отсчетов зависит от требуемой точности восстановления исходного сообщения и вида применяемой интерполяции.

Для оценки необходимого интервала дискретизации в общем случае можно использовать соотношение

$$\Delta t_k \approx (\delta_0)^{1/(k+1)} \Delta t_{\text{пр}} = \frac{1}{2F_{\text{в}}} (\delta_0)^{1/(k+1)}, \qquad (2.31)$$

где k — порядок интерполяции.

Можно показать, что величина ошибки равномерного приближения, определяемая выражением (2.30), близка к величине сред-

неквадратичной ошибки дискретизации, определяемой выражением (2.19), т. е.

$$\delta_0 \approx \delta_д \approx \sqrt{2\Delta E_x/E_x}. \qquad (2.32)$$

Это обстоятельство позволяет указать следующую методику инженерного расчета необходимого интервала дискретизации при выбранном виде интерполяции:

1) по требуемой точности дискретизации δ_0 или $\delta_д$ и выражению $G_x(\Omega)$ для энергетического спектра случайного процесса, описывающего рассматриваемый класс непрерывных сообщений, определяют верхнюю граничную частоту спектра $\Omega_в$;

2) по частоте $\Omega_в$ находят предельный интервал дискретизации, соответствующий теореме Котельникова;

3) для выбранного вида интерполяции рассчитывают необходимый интервал дискретизации, используя приведенные выше формулы.

Необходимо отметить, что изложенная методика приемлема лишь для сообщений с быстро спадающими спектрами. Для сообщений со сравнительно медленно спадающими спектрами расчеты по изложенной методике дают уменьшенные значения интервалов дискретизации, особенно для способов интерполяции высокого порядка (выше второго). Сказанное означает, что для сообщений со сравнительно медленно спадающими спектрами нецелесообразно применять способы интерполяции высокого порядка (второго и выше), так как такие способы малоэффективны *.

Исследованию вопросов дискретизации непрерывных сообщений с учетом различных интерполяционных способов восстановления посвящено значительное число работ. Однако многие вопросы еще не решены. В частности, до сих пор не получены строгие соотношения для определения необходимых интервалов дискретизации с учетом конечной скорости спадания энергетического спектра сообщения и вида интерполяции. Большой интерес представляет отыскание оптимальных способов восстановления и определение их параметров с учетом допустимой избыточности отсчетов и действия помех в канале связи.

С некоторыми вопросами теории и расчета интерполяционных способов восстановления можно ознакомиться в [2, 3, 16].

Помимо изложенных выше интерполяционных способов восстановления, на практике находит применение метод непосредственного восстановления исходного непрерывного сообщения по его отсчетам с помощью аналоговых фильтров нижних частот или регистрирующих устройств передаточные функции которых подобны характеристикам указанных фильтров.

---

* Деление спектров на быстро и медленно спадающие, до некоторой степени условно. В инженерных приложениях можно принять, что спектр является медленно спадающим, если уменьшение спектральной плотности происходит медленнее, чем изменяется величина $1/\Omega^3$—$1/\Omega^4$.

2.3. ДИСКРЕТИЗАЦИЯ НЕПРЕРЫВНЫХ СООБЩЕНИЙ ПО УРОВНЮ (КВАНТОВАНИЕ)

Как указывалось ранее, каждое непрерывное сообщение нужно рассматривать как одну из реализаций, принадлежащих к бесконечному множеству реализаций некоторого случайного процесса. Мгновенные значения этого процесса могут быть любыми в пределах некоторого диапазона $[x_{max}, x_{min}]$, определяемого физическими ограничениями, характерными для реальных сообщений. Этот диапазон принято называть непрерывной шкалой мгновенных значений сообщений.

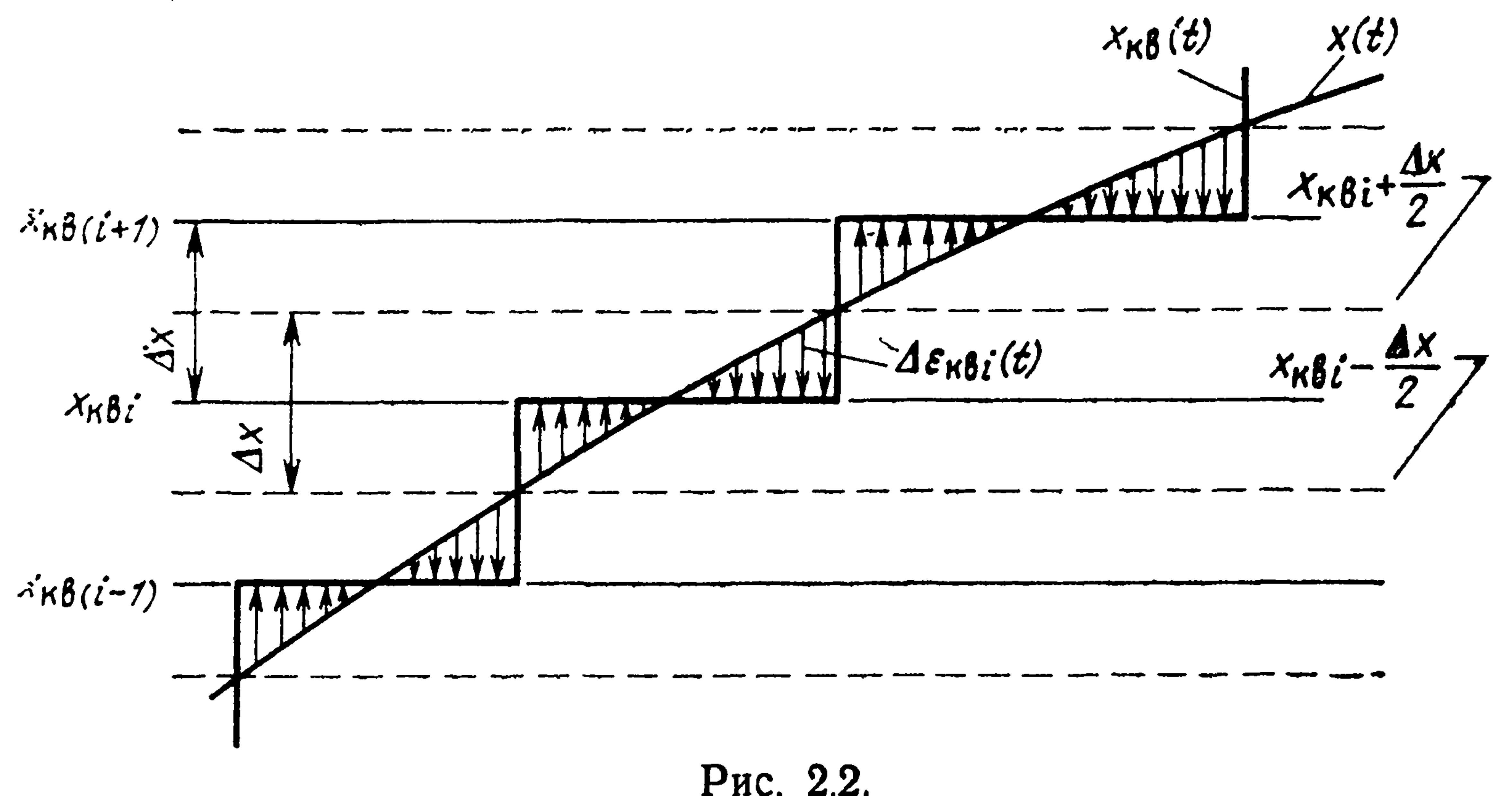


Рис. 2.2.

В реальных условиях из-за влияния внутренних шумов, различных помех, аппаратурной нестабильности и конечной разрешающей способности получателей передача сообщений всегда сопровождается ошибкой. Это означает, что сообщения, достаточно близкие по форме («похожие» друг на друга), различить при приеме невозможно. Появляется как бы некоторая зона неразличимости, в пределах которой нельзя установить истинное значение переданного сообщения. Следовательно, вовсе не обязательно стремиться к тому, чтобы передавать сообщения точнее, чем это позволяют условия передачи и характеристики системы связи. Можно заранее учесть все или часть факторов, вызывающих искажения и ошибки при передаче сообщений, и в соответствии с этим преобразовать передаваемые сообщения так, чтобы эти факторы уже никакого влияния на передачу сообщений не оказывали. Естественно, что такое преобразование будет связано с некоторой ошибкой, которая вводится сознательно в передающей части системы связи. Установим основные особенности такого преобразования.

Разобьем непрерывную шкалу мгновенных значений передаваемых сообщений на конечное число частей (квантов). Полученная при этом дискретная шкала называется шкалой уровней квантования, а деления шкалы называются уровнями квантования. Интервал

между соседними уровнями квантования называется шагом квантования. Различают квантование равномерное, когда шаг квантования Δx постоянен по всей шкале, и неравномерное, когда шаг квантования изменяется от уровня к уровню по некоторому правилу. Равномерное квантование технически реализуется значительно проще, и поэтому оно получило наибольшее распространение.

В соответствии с введенной шкалой квантования можно произвести замену непрерывного сообщения квантованным по уровню, т. е. таким, мгновенные значения которого могут принимать лишь конечное число дискретных значений, определяемых уровнями квантования. Для иллюстрации сказанного рассмотрим рис. 22, на котором показана только часть шкалы квантования (три уровня) и соответствующее этой части квантование. На этом рисунке обозначено: $x_{\text{кв}(i-1)}$, $x_{\text{кв}i}$, $x_{\text{кв}(i+1)}$ — уровни квантования с номерами $(i-1)$, i и $(i+1)$; Δx — шаг квантования; $\Delta\varepsilon_{\text{кв}i}(t)$ — ошибка квантования.

Правило квантования очень простое: мгновенные значения сообщения каждый раз относят к ближайшему уровню. Нетрудно видеть, что при таком правиле каждый уровень квантования должен находиться в середине зоны неразличимости, равной шагу квантования.

В результате квантования непрерывное сообщение $x(t)$ заменяется сообщением $x_{\text{кв}}(t)$, которое имеет ступенчатую форму и может принимать только конечное число различных мгновенных значений, равное числу уровней квантования. Естественно, что при такой замене появляется ошибка, которую называют ошибкой квантования.

Сказанное можно выразить соотношением

$$x_{\text{кв}}(t)=x(t)+\varepsilon_{\text{кв}}(t). \qquad (2.33)$$

Ошибка квантования представляет собой случайный процесс. Поэтому ее часто называют шумом квантования Количественно влияние шума квантования обычно оценивают величиной относительной среднеквадратичной ошибки (СКО квантования)

$$\delta_{\text{кв}}=\left[\overline{\varepsilon^2_{\text{кв}}(t)}/\overline{x^2(t)}\right]^{1/2}, \qquad (2.34)$$

где $\overline{\varepsilon^2_{\text{кв}}(t)}$ и $\overline{x^2(t)}$ — средние квадраты (удельные мощности) шума квантования и квантуемого сообщения соответственно

Считая процессы $\varepsilon_{\text{кв}}(t)$ и $x(t)$ эргодическими, можно показать, что при равномерном квантовании

$$\delta_{\text{кв}}=\frac{1}{2\sqrt{3}}\,\frac{A_x}{\sqrt{\overline{x^2(t)}}}\,\frac{1}{L_{\text{кв}}}=\frac{k_x}{L_{\text{кв}}}, \qquad (2.35)$$

где A_x — полная шкала мгновенных значений процесса $x(t)$; $L_{\text{кв}}$ — число уровней квантования; k_x — коэффициент, зависящий от закона распределения мгновенных значений процесса $x(t)$.

Исследования, проведенные для ряда распределений (равномерного, арксинуса и нормального), показывают, что величина k_x заключена в пределах $0{,}8 \leqslant k_x \leqslant 1{,}6$. При инженерных расчетах обычно принимают $k_x=1$.

Более подробно с вопросами квантования непрерывных сообщений можно ознакомиться в работах [14—18].

2.4. ПРЕОБРАЗОВАНИЕ ДИСКРЕТИЗИРОВАННЫХ СООБЩЕНИЙ В ЦИФРОВУЮ ФОРМУ

В результате дискретизации по времени и по уровню непрерывный случайный процесс, являющийся математической моделью непрерывных сообщений, заменяется дискретным случайным процессом. Такой процесс представляет собой последовательность отсчетов, каждый из которых может принять любое мгновенное значение из конечного множества возможных целых значений, соответствующих выбранным уровням квантования.

Таким образом, величина каждого отсчета представляет собой случайное число, которое может принять одно из $L_{кв}$ возможных целых значений. Следовательно, передача непрерывных сообщений заменяется передачей последовательности чисел. Эти числа можно выразить в удобной системе счисления и передать по линии связи в виде последовательности некоторых кодовых комбинаций.

Для того чтобы представить дискретизированные сообщения в цифровой форме, необходимо выбрать систему счисления. Эта система должна быть удобной с точки зрения реализации нужных кодовых комбинаций.

При выборе основания системы счисления в первую очередь необходимо учитывать простоту, экономичность и удобство реализации цифрового представления информации. Пусть для представления чисел используется система с основанием n и числом разрядов m. Тогда количество элементов, используемых в данной системе при записи чисел, равно

$$v = mn. \qquad (2.36)$$

Максимальное число, которое может быть записано в такой системе счисления, равно

$$\max N(n) = n^m - 1. \qquad (2.37)$$

Из выражения (2.37) следует

$$m = \ln[\max N(n) + 1]/\ln n. \qquad (2.38)$$

С учетом (2.38) выражение (2.36) принимает вид

$$v = \frac{n}{\ln n} \ln[\max N(n) + 1]. \qquad (2.39)$$

Найдем условие, при котором для записи числа $\max N$ потребуется минимальная величина v:

$$\frac{dv}{dn}=\ln[\max N+1]\frac{\ln n-1}{(\ln n)^2}=0. \qquad (2.40)$$

Нетрудно убедиться, что это условие минимума величины v.

Из условия (2.40) следует, что оптимальной системой счисления, при которой требуется минимальное количество элементов записи чисел, является система счисления с основанием

$$n=n_{опт}=e\approx 2{,}72. \qquad (2.41)$$

Практически наиболее близка к оптимальной троичная система ($n=3$). Цифры троичной системы (0, 1, 2) можно реализовать в виде символов (+1, 0, —1), которые представляют три различных состояния (например, импульс положительной полярности, пауза и импульс отрицательной полярности) некоторого элемента троичной логики. Однако такие элементы уступают в простоте, надежности и быстродействии элементам двоичной логики. Поэтому в технических приложениях наибольшее применение находит двоичная система счисления ($n=2$). Реализация двух цифр такой системы (0 и 1) наиболее проста.

В настоящее время известно много простых и надежных устройств, которые могут принимать два устойчивых состояния, соответствующих двум цифрам двоичной системы счисления (различные реле, триггеры и т. п.). Различие в требуемой величине v для разных систем счисления иллюстрируется табл. 2.1.

Таблица 2.1

n	2	2,72	3	4	8	10	12
$n/\ln n$	2,88	2,72	2,735	2,88	3,85	4,3	4,83
v/v_{opt}	1,06	1	1,006	1,06	1,42	1,58	1,77

Из рассмотрения таблицы следует, что для систем счисления с основанием $n=2$—4 отличие от оптимального случая незначительно. Поэтому выбор основания

счисления должен в первую очередь определяться практической стороной дела. Надо применять ту систему счисления, которая наиболее целесообразна в данных конкретных условиях. В дальнейшем главное внимание будет уделяться цифровому представлению дискретной информации, основанному на двоичной системе счисления.

Итак, каждому целому числу, а следовательно, каждому уровню квантования можно поставить в соответствие двоичное число, которое реализуется в виде кодовой комбинации, составленной из некоторых символов. Естественно, что число кодовых комбинаций (кодовых групп) $N_к$ не должно быть меньше числа уровней квантования ($N_к \geqslant L_{кв}$). Аналогично поступают и при переводе любой другой информации дискретного характера (например, буквенной, буквенно-цифровой и т. п.) в цифровую форму. Система соответствий между элементами дискретных сообщений (числами, буквами, знаками и т. п.) и кодовыми комбинациями, с помощью которых они могут быть представлены и переданы по каналу связи, называется кодом *.

Совокупность кодовых комбинаций, которая используется для представления заданного количества дискретных сообщений, называется кодовым словарем. Однозначное соответствие между элементами дискретных сообщений и кодовыми комбинациями словаря можно обеспечить различными видами кодов. Каждый конкретный код характеризуется определенными правилами, по которым образуются комбинации кодового словаря.

Одним из главных параметров кода является его основание, основание кода соответствует основанию выбранной системы счисления. Оно определяет число различных символов, с помощью которых образуются кодовые комбинации. Эти символы часто называют элементами кода, а совокупность используемых элементов, число которых равно основанию счисления, — алфавитом кода.

Коды, у которых основание равно двум, называются двоичными или бинарными. Коды с основанием больше двух принято называть многоосновными.

---

* Термин «код» происходит от французского слова code, которое, в свою очередь, произошло от латинского слова codex — «кодекс», т. е. книга, содержащая свод законов, систему правил и т. п.

Далее будем рассматривать в основном двоичные коды, комбинации которых строятся с использованием только двух элементов x_1 и x_2. В литературе встречаются различные условные обозначения элементов двоичного кода. Наиболее употребительные из них приведены в табл. 2.2.

Таблица 2.2

x_1	1	+1	+	Z	M	▨
x_2	0	-1	–	A	S	▭

Обозначения $x_1=Z$, $x_2=A$ рекомендованы МККТТ *. Обозначения $x_1=M$ («Mark» — отметка, знак), $x_2=S$ («Space» — расстояние, промежуток) часто применяются в иностранной литературе.

При реализации двоичных кодов необходимо представлять их элементы (символы) в форме, удобной для выполнения последующих операций и передачи образованных кодовых последовательностей по линиям связи. В настоящее время известно большое число различных форм представления элементов двоичных кодов. Часть из них приведена на рис. 2.3. Выбор той или иной формы представления зависит от многих обстоятельств и условий, в частности, от вида используемой манипуляции несущей сигнала, требований к точности символьной синхронизации в приемнике, допустимой полосы пропускания приемника и скорости передачи и т. д. Поэтому дать какие-либо рецепты по этому поводу невозможно — это творческая инженерная задача. Ниже будут лишь приведены некоторые соображения, которые полезно иметь в виду при решении указанной задачи.

Для рассмотренных форм представления элементов двоичных кодов известен ряд способов формирования этих элементов. Все эти способы можно разделить на две группы:

1. Формирование элементов кода методом возврата

* МККТТ — международный консультативный комитет по телефонии и телеграфии — организация, занимающаяся разработкой различных рекомендаций и предложений в области телефонии, телеграфии и передачи данных.

№ п/п	Форма символов		Название символов
	$x_1=1$	$x_2=0$	
1	U_1, 0, τ_0, t	0, t	Униполярные с пассивной паузой
2	U_1, 0, τ_0, t	0, τ_0, $-U_1$, t	Полярные с активной паузой
3	U_1, $\tau_0/2$, t	0, t	Униполярные с пассивной паузой и расщепленной фазой
4	U_1, 0, $\tau_0/2$, t, U_1	$\tau_0/2$, U_1, t, $-U_1$	Полярные с расщепленной фазой
5	U_1, 0, $\tau_0/2$, t	U_1, 0, $\tau_0/2$, t	Униполярные с расщепленной фазой
6	U_1, 0, $\tau_0/2$, t	t, 0, $\tau_0/2$, $-U_1$	Полярные с расщепленной фазой

Рис. 2.3.

к нулевому уровню. Этот метод сокращенно обозначают ВН (возврат к нулю). В иностранной литературе он известен как метод RZ (return-to-zero).

2. Формирование элементов кода методом невозвращения к нулевому уровню. Этот метод сокращенно обозначают БВН (без возврата к нулю). В иностранной литературе он известен как метод NRZ (no-return-to-zero).

Особенностью метода ВН является то, что каждый элемент кода начинает формироваться от некоторого нулевого (исходного) уровня и заканчивается также на этом уровне. В методе БВН для получения элементов кода используются два уровня — положительный и отрицательный, расположенные симметрично относительно исходного нулевого уровня. При этом в процессе формирования исходный нулевой уровень не используется. Напряжение на выходе устройства формирования кода по методу БВН может принять значения нулевого уровня только при отсутствии информационных симво-

лов. В зависимости от формы применяемых символов методы ВН и БВН позволяют получить значительное число разновидностей двоичных кодов. Эти разновидности кодов можно разделить на две категории:

1) Коды, у которых в формируемой последовательности изменение уровня или фазы элемента (для униполярных элементов с расщепленной фазой) происходит при каждом переходе от символа 0 к символу 1 или от 1 к 0. Назовем их абсолютными кодами.

2) Коды, у которых изменение уровня или фазы элемента происходит лишь при появлении символа 1, а при появлении символа 0 сохраняется значение уровня или фазы предыдущего элемента. Возможен и другой случай, когда соответствующее изменение происходит только при появлении символа 0, а появление символа 1 никаких изменений не вызывает. Такие коды назовем относительными.

В иностранной литературе абсолютные коды иногда обозначают буквой L, а относительные — M или S в зависимости от того, появление какого символа вызывает изменение уровня или фазы элемента.

Примеры формирования различных кодов, образованных по методам ВН и БВН, приведены на рис. 2.4. Рассмотрение этих примеров показывает, что, выбирая соответствующий метод формирования кода, можно получить последовательности, эквивалентные в информационном смысле, но существенно различающиеся по своим статистическим свойствам. В частности, различными получаются энергетические спектры сформированных последовательностей. В некоторых из них (например № 3 и 5 на рис. 2.4) отсутствует постоянная составляющая и существенно ослаблена низкочастотная часть спектра. Используя последовательности с такими спектрами, можно получить систему с повышенной помехоустойчивостью к низкочастотным шумам и вибрациям. В ряде кодов длительность импульсов и пауз в последовательности не превышает удвоенной длительности информационного символа. Это позволяет легче выделить частоту следования символов, необходимую для обеспечения символьной синхронизации в приемнике.

Из рассмотренных выше кодов в настоящее время довольно широкое применение находят коды с расщепленной фазой элемента. При этом метод БВН предпочтительнее метода ВН. Некоторые дополнительные све-

дения по вопросам формирования кодов можно найти в [19—21]. Расчет спектров различных кодовых последовательностей может быть проведен по методике, изложенной в [5, 19].

Каждая сформированная кодовая комбинация состоит из некоторого количества элементов, которое определяет ее значность. Если все кодовые комбинации содержат одинаковое количество элементов, код называется

№ п/п	Вид последовательности символов								Метод формирования кода
	1	0	1	1	0	0	1	1	
1	0 … t								ВН Символы униполярные с пассивной паузой
2	0 … t								ВН Символы униполярные с расщепленной фазой
3	0 … t								ВН Символы полярные с расщепленной фазой
4	0 … t								БВН Символы полярные с активной паузой
5	0 … t								БВН Символы полярные с расщепленной фазой (№ 4 на рис. 2.3)
6	0 … t								ВН Относительный код S
7	0 … t								ВН Относительный код M

Рис. 2.4.

равномерным. Если кодовые комбинации содержат различное количество элементов, то код называется неравномерным. Примером неравномерного кода является телеграфный код Морзе, в котором буквам, используемым чаще, присваиваются короткие кодовые комбинации, а редко используемым буквам — длинные. В равномерном коде длина всех комбинаций одинакова, а сле-

довательно, одинакова и значность. Очевидно, что равномерные коды при передаче неравновероятных дискретных сообщений менее эффективны, чем неравномерные. Однако техническая реализация равномерных кодов значительно проще, чем неравномерных, и поэтому они получили в современных системах связи наиболее широкое применение.

Пусть имеется N различных элементов дискретных сообщений (чисел, букв, команд и т. п.), которым нужно сопоставить кодовый словарь равномерного кода с основанием n и значностью m. Количество кодовых комбинаций такого кода должно удовлетворять условию

$$N_к = n^m \geqslant N. \tag{2.42}$$

Если $N_к = N$, все комбинации будут использованы. Однако такой случай довольно редок, так как N обычно не бывает кратным степеням выбранной системы счисления. Поэтому, как правило, должно выполняться условие $N_к > N$. Следовательно, некоторая часть кодовых комбинаций не будет использована и оказывается лишней (избыточной). Код, для которого выполняется условие

$$n^{m-1} < N \leqslant n^m, \tag{2.43}$$

называется простым, или первичным кодом. Характерной особенностью такого кода является то, что в нем нельзя уменьшить значность кодовых комбинаций, поскольку при уменьшении значности кода на единицу число комбинаций становится меньше числа дискретных сообщений. Нетрудно видеть, что условие (2.43) эквивалентно условию

$$1/n < N/N_к \leqslant 1, \tag{2.44}$$

которое для равномерных первичных двоичных кодов принимает вид

$$0{,}5 < N/N_к \leqslant 1. \tag{2.45}$$

Из этого условия следует, что в двоичном первичном коде неиспользуемая часть кодовых комбинаций всегда меньше половины общего числа кодируемых дискретных величин. Если первичный код применяется для передачи аналоговых сообщений в цифровой форме, то, как правило, используются все кодовые комбинации, что автоматически обеспечивается при работе преобразователя «аналог—цифра».

Первичные коды служат для экономного представления заданного множества дискретных величин кодовыми комбинациями, иными словами, основное назначение первичных кодов — представлять дискретную информацию в цифровой форме. Помимо первичных кодов при передаче информации находит применение и другой класс кодов — корректирующие коды. Такие коды позволяют обнаруживать и исправлять ошибки, возникающие при передаче дискретной информации. С особенностями корректирующих кодов мы ознакомимся в гл. 6.

2.5. ОСНОВНЫЕ ВИДЫ ПЕРВИЧНЫХ КОДОВ И ИХ ОСОБЕННОСТИ

Все первичные коды в зависимости от того, присваивается или не присваивается каждому разряду кодовой комбинации некоторый весовой коэффициент, разделяются на взвешенные и невзвешенные. Так как каждый разряд в кодовой комбинации занимает определенную позицию, взвешенные коды иногда называют позиционными, а невзвешенные — непозиционными. Наиболее часто значения весовых коэффициентов берут кратными целым степеням основания кода.

Основной формой представления целых чисел и других дискретных величин в цифровых системах связи является истинная двоичная форма, а соответствующий ей код называется естественным или натуральным двоичным кодом. Это наиболее распространенный вид взвешенных равномерных первичных кодов.

В двоичном натуральном коде веса разрядов определяются последовательными степенями числа 2. Если комбинации кода состоят из m разрядов, то веса в порядке убывания старшинства разрядов имеют значения 2^{m-1}, 2^{m-2}, ..., 2^1, 2^0. Двоичный натуральный код прост в реализации и удобен при выполнении вычислительных операций в ЭВМ и других автоматических устройствах, так как обеспечивает простоту логики работы таких устройств и их высокую надежность.

Однако такому коду присущ существенный недостаток, связанный с явлением, получившим название «неоднозначность отсчета». Иногда это явление называют «неопределенностью границы считывания». Суть этого явления состоит в том, что при считывании сформиро-

ванной кодовой комбинации в кодирующем устройстве может возникать ложная комбинация, которая существенно отличается от истинной. Рассмотрим указанное явление подробнее.

При переходе от одного дискретного значения, выраженного некоторой комбинацией натурального двоичного кода, к другому может возникать ситуация, когда в новой комбинации изменяются значения сразу нескольких разрядов (позиций). Например, при переходе от числа 7, которому соответствует комбинация 0111, к числу 8, которому соответствует комбинация 1000, произошло изменение состояний во всех разрядах (позициях) по сравнению с предыдущей. Таким образом, при изменении предыдущей комбинации на единицу младшего разряда произошло изменение во всех позициях вновь образованной комбинации. Для m-разрядного двоичного кода аналогичное положение возникает каждый раз, когда происходит переход от числа $2^{i-1}-1$ к числу 2^{i-1}, где $i=2, 3, \ldots, m$. В этом случае предыдущая и последующая кодовые комбинации отличаются в i разрядах (позициях). Наибольшее отличие получается при переходе от числа $2^{m-1}-1$ к числу 2^{m-1} или наоборот. Соответствующие кодовые комбинации различаются во всех m позициях.

Указанная особенность означает, что при некоторых способах первичного кодирования должна происходить смена состояний всех ячеек блока памяти, входящего в состав кодирующего устройства. Такое положение может возникать при первичном кодировании, основанном, например, на счетно-импульсном методе, или при кодировании по методу считывания. Аналогичное положение возможно также при некоторых преобразованиях дискретных сообщений, когда требуется запоминать кодовые комбинации на некоторое время, а затем выводить (считывать) их из устройств памяти. Если вывод информации из блока памяти (считывание) начинается раньше, чем успевают установиться все состояния ячеек памяти, может быть считана любая из промежуточных комбинаций и ошибка достигнет большой величины. Чтобы избежать этого, нужно начинать считывание после того, как перенос во всех разрядах кодовой комбинации закончится, т. е. все состояния ячеек памяти установятся. Стремление повысить быстродействие кодирующего устройства может привести к появлению ошибок считы-

вания из-за временного рассогласования между считывающими импульсами и соответствующими позициями кодовых групп. Это рассогласование появляется в результате нестабильности аппаратурных характеристик. Если временное рассогласование $\Delta\tau_p$ превысит половину длительности позиции τ_0, соответствующей старшему разряду, то вместо элемента кода с весом 2^{m-1} будет считан элемент с весом 2^{m-2} и произойдет значительная ошибка.

Известны способы, которые позволяют устранить неоднозначность считывания при использовании двоичных натуральных кодов. К таким способам относятся фиксация момента считывания; запрещение опроса ячеек памяти во время переноса разрядов; применение преобразования по методу двойных проб; считывание натурального двоичного кода по способу V-развертки [22].

Двоичному натуральному коду присущ еще один существенный недостаток. Как уже указывалось, этот код взвешенный. При передаче информации по каналам связи под действием помех отдельные элементы кода могут так исказиться, что окажутся принятыми неверно. Например, вместо посланного элемента 0 будет принят элемент 1 или наоборот. При использовании взвешенного кода небезразлично, какой элемент в данной кодовой группе принят неверно. Если неверно принят элемент, соответствующий позиции старшего разряда, ошибка в передаче числа значительно больше, чем при неверном приеме элемента младшего разряда. С этой точки зрения лучше применять невзвешенный код, у которого ошибки, вызванные помехами, были бы одинаковыми для любого разряда.

Широкое развитие автоматизированных средств передачи информации, применение вычислительных машин при обработке информации и передаче данных заставили в ряде случаев отказаться от применения в цифровых системах связи натурального кода из-за свойственных ему недостатков. В последние 10—15 лет для представления аналоговых сообщений в цифровой форме широкое применение получили рефлексные коды *. Рассмотрим некоторые особенности таких кодов.

---

* Термин «рефлексные» коды происходит от латинского слова reflexio — отражение. Такое название, как мы убедимся ниже, хорошо соответствует их особенностям.

Рефлексные коды относятся к группе невзвешенных кодов. В незвешенных кодах позициям (разрядам) кодовой комбинации не приписывают определенных весов. Вес имеет лишь вся кодовая комбинация в совокупности. Характерной особенностью рефлексных кодов является то, что любые две соседние m-значные кодовые комбинации отличаются друг от друга лишь в одной позиции. Наиболее часто применяемым двоичным рефлексным кодом является код Грея [18, 22]. Для полу-

Таблица 2.3

Десятичное число	Натуральный двоичный код	Код Грея
0	0 0 0 0	0 0 0 0
1	0 0 0 1	0 0 0 1
2	0 0 1 0	0 0 1 1
3	0 0 1 1	0 0 1 0
4	0 1 0 0	0 1 1 0
5	0 1 0 1	0 1 1 1
6	0 1 1 0	0 1 0 1
7	0 1 1 1	0 1 0 0
8	1 0 0 0	1 1 0 0
9	1 0 0 1	1 1 0 1
10	1 0 1 0	1 1 1 1
11	1 0 1 1	1 1 1 0
12	1 1 0 0	1 0 1 0
13	1 1 0 1	1 0 1 1
14	1 1 1 0	1 0 0 1
15	1 1 1 1	1 0 0 0

чения комбинации в коде Грея можно воспользоваться следующим простым правилом: кодовую комбинацию натурального кода складывают по модулю 2 с такой же комбинацией, сдвинутой на один разряд вправо, при этом младший разряд сдвинутой комбинации отбрасывается *.

Рассмотрим особенности кода Грея на примере. В табл. 2.3 представлен ряд чисел, выраженных комбинациями натурального двоичного кода и кода Грея, полученного по изложенному выше правилу. Для про-

* Сложение по модулю 2 выполняется по правилам $1 \oplus 1 = 0$; $1 \oplus 0 = 1$; $0 \oplus 1 = 1$; $0 \oplus 0 = 0$.

стоты рассмотрения взято 16 дискретных значений (включая нуль), цифровое представление которых требует всего четыре разряда двоичного кода.

Рассмотрение таблицы позволяет установить ряд характерных особенностей кода Грея:

1. Каждая последующая комбинация всегда отличается от предыдущей только в одной позиции (в одном разряде); как уже указывалось выше, для натурального двоичного кода это свойство не соблюдается.

2. Смена значений элементов в каждом разряде (1 на 0 или 0 на 1) при переходе от комбинации к комбинации кода Грея происходит вдвое реже, чем в натуральном двоичном коде.

Поясним сказанное. В натуральном двоичном коде смена элемента первого (младшего) разряда происходит с чередованием элементов 0 — 1 — 0 — 1..., второго разряда с чередованием 00 — 11 — 00 — 11..., третьего — с чередованием 0000 — 1111 — 0000... В коде Грея соответственно имеем следующие чередования элементов: для первого разряда 11 — 00 — 11..., для второго 0000 — 1111 — 0000... Это свойство кода Грея позволяет при том же быстродействии схемы кодирования получать точность кодирования выше по сравнению с натуральным двоичным кодом.

3. При сложении двух соседних комбинаций кода Грея по модулю 2 (mod 2) число единиц равно числу разрядов минус три, т. е. в нашем случае имеем одну единицу. В общем случае, для m-значного кода число единиц равно m—3. Это свойство кода Грея иногда можно использовать для проверки правильности принятых комбинаций.

4. В коде Грея можно выделить оси симметрии (оси «отражения»), относительно которых наблюдается идентичность элементов в некоторых разрядах. Так, например, имеет место симметрия относительно оси, проведенной между числами 7 и 8. В комбинациях, симметричных относительно этой оси, идентичны три символа младших разрядов (см. табл. 2.3). Относительно осей 1 — 2, 5 — 6, 9 — 10, 11 — 12 имеется симметрия в одном символе младшего разряда, а относительно осей 3 — 4 и 11 — 12 — в двух символах. Ось симметрии, которая проходит в m-значном коде Грея между комбинациями, соответствующими уровням (числам) 2^{m-1}—1 и 2^{m-1}, называется главной осью симметрии. Относительно нее

имеется идентичность символов в $(m-1)$ разрядах симметричных кодовых комбинаций. Отмеченная особенность и послужила основанием для введения термина «рефлексный» (или отраженный) код.

По сравнению с натуральным кодом рефлексный код Грея позволяет уменьшить ошибки неоднозначности считывания, а также ошибки, вызванные помехами при передаче информации по каналам связи. Если, например, неправильное воспроизведение символа старшего разряда комбинации натурального кода, соответствующего числу 2^{m-1}, приводит к значительной ошибке, то для кода Грея эта ошибка составит величину одного младшего разряда. В частности, в соответствии с табл. 2.3 для числа 8 комбинация 1000 натурального двоичного кода перейдет в нулевую, а комбинация 1100 кода Грея — в ближайшую комбинацию 0100, соответствующую числу 7.

Рассмотренные свойства кода Грея показывают, что такой код удобен для аналого-цифрового преобразования различных непрерывных сообщений и их передачи по каналам связи. Недостатком кода Грея и других рефлексных кодов является то, что эти коды невзвешенные. Информацию, представленную такими кодами, трудно обрабатывать с помощью ЭВМ. Декодирование таких кодов в приемных устройствах цифровых систем связи также связано с большими трудностями из-за существенного усложнения декодирующих устройств (преобразователей «аналог — цифра»). Поэтому перед вводом в ЭВМ или перед декодированием в приемных устройствах рефлексный код обычно преобразуют в натуральный двоичный код, который удобен для ЭВМ, а также сравнительно легко декодируется в приемных устройствах.

Преобразование кода Грея в натуральный двоичный код выполняется по следующему правилу: если преобразование необходимо начинать со старших разрядов кода Грея, символ старшего разряда записывается без изменения, каждый следующий символ кода Грея нужно инвертировать, если в натуральном коде перед этим был получен символ 1, или оставить без изменения, если в натуральном коде был получен символ 0. Более подробно ознакомиться с вопросами преобразования кодов можно в работах [22, 23].

Кроме рассмотренных выше двух важных видов первичных кодов, известен ряд других первичных кодов,

которые также находят довольно широкое применение. К таким кодам относятся рекуррентные циклические и двоично-десятичные коды.

Рекуррентные циклические коды являются одной из разновидностей невзвешенных равномерных кодов. Они могут быть получены из последовательности, содержащей 2^m-1 символов 0 и 1, циклическим формированием m-символьных кодовых комбинаций. Каждая последующая комбинация получается из предыдущей сдвигом вдоль последовательности на одну позицию (разряд) в том или другом направлении. Если брать достаточно продолжительные последовательности и изменять начало и направление обхода, можно получить большое число различных комбинаций. Это свойство рекуррентных кодов можно использовать при создании крипто- и имитостойких систем связи.

Остановимся кратко на особенностях двоично-десятичных кодов. Такие коды применяются, когда числа желательно представить в цифровой двоичной форме, удобной для потребителя. Каждая десятичная цифра в таком коде представлена комбинацией натурального двоичного кода. Для этого достаточно использовать любые 10 комбинаций двоичного 4-разрядного натурального кода. В настоящее время наиболее часто применяются первые 10 комбинаций (включая и нулевую), которые приведены в табл. 2.4.

Таблица 2.4

Десятичные цифры	0	1	2	3	4	5	6	7	8	9
Натуральный код	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001

Число, которое необходимо представить в двоично-десятичном коде, кодируется цифра за цифрой. Каждая из этих цифр в соответствии с табл. 2.4 представляет собой кодовую комбинацию, состоящую из четырех символов двоичного натурального кода. Такая комбинация называется тетрадой.

Например, число 1975, выраженное в двоично-десятичном коде, представляет собой последовательность тетрад в виде 0001 1001 0111 0101.

Кодирование десятичных чисел тетрадами уступает по экономичности кодированию натуральным двоичным кодом. Если десятичное число содержит m разрядов, то при двоично-десятичном кодировании потребуется $4m$ двоичных символов, а при натуральном двоичном кодировании — $m \log_2 10 \approx 3{,}3m$ двоичных символов.

Кроме того, при двоично-десятичном кодировании необходимо вводить разграничительные интервалы между тетрадами, что приводит к уменьшению быстродействия аппаратуры и скорости передачи информации.

Известен ряд разновидностей двоично-десятичных кодов, удобных для обработки числовой информации на вычислительных машинах. Особенности таких кодов детально рассмотрены в работах [22, 24].

В настоящее время в различных областях приложений радиотехники уделяется большое внимание вопросам стандартизации. Большая работа проводится по стандартизации и регламентированию в области различных систем передачи цифровой информации [22]. В международном плане эту работу координируют организации МОС, МККР и МККТТ. Безусловно, что при разработке новых систем передачи информации, средств ее обработки и т. п. необходимо строго придерживаться уже разработанных стандартов и предложений. Однако, к сожалению, стандартизация в некоторых областях применения систем передачи информации еще не достигла высокого уровня. В частности, первичные коды в ряде систем до сих пор не стандартизированы. Исключение составляют телеграфные системы и системы передачи данных. В этих системах применяются только стандартные первичные коды.

2.6. КРАТКИЕ СВЕДЕНИЯ О ПЕРВИЧНЫХ СТАНДАРТНЫХ КОДАХ

В телеграфии двоичные коды для передачи буквенных текстов, цифр и других дискретных сообщений широко применяются уже более ста лет. Многие годы широкое применение в телеграфной связи находил простой двоичный пятизначный код, предложенный французским инженером Ж. Бодо в 1874 г. Буквы, цифры и некоторые служебные знаки (точки, запятые, скобки и т. п.) передавались сочетаниями из символов 1 и 0 на

пяти позициях равномерного кода. Поскольку пятизначный двоичный код позволяет получить только 32 кодовые комбинации, а число необходимых кодовых комбинаций для передачи кодом Бодо различных букв, цифр и знаков равно 56 (см. [22]), то при использовании кода Бодо применяется двухрегистровый режим работы. Суть этого режима состоит в том, что все дискретные величины (буквы, цифры, знаки) разделяются на две равные группы (два регистра). В нашем случае в каждой группе (регистре) окажется по 28 дискретных величин. Каждому регистру приписывается специально выделяемая кодовая комбинация, которая является его признаком (адресом).

Комбинация, являющаяся адресом данного регистра, всегда предшествует передаче величин, принадлежащих этому регистру. При переходе от одного регистра к другому производится смена адреса и те же самые кодовые комбинации используются теперь для передачи величин, входящих в состав второго регистра.

В конце XIX века американский инженер Дж. Муррей предложил двоичный пятизначный код, который отличался от кода Бодо тем, что для наиболее часто встречающихся букв и знаков использовались кодовые комбинации с наименьшим числом ненулевых разрядов. Такой код позволил уменьшить нагрузку на телеграфную аппаратуру и повысить надежность ее работы.

Коды Бодо и Муррея многие годы широко применялись на телеграфных линиях связи. Кроме того, находили применение и некоторые другие разновидности двоичных пятизначных кодов. В условиях быстрого развития телеграфной связи применение различных кодов препятствовало унификации телеграфной аппаратуры и затрудняло развитие международных телеграфных линий связи. В конце 20-х годов XX века назрела необходимость в стандартизации телеграфного кода.

В 1931 г. МККТТ принял стандартный код № 1, в основу которого положен код Бодо. В 1932 г. был принят стандартный код № 2, в основу которого положен код Муррея.

Код № 2 содержит все основные знаки, которые установлены МККТТ как необходимые для обеспечения телеграфного обмена. Для телеграфных аппаратов с русско-латинским алфавитом был разработан вариант международного кода № 2 (см. табл. 2.5). Таблица этого

кода составлена так, чтобы на регистрах «буквы латинские» и «цифры» она совпадала с международным стандартом. Для русского алфавита введен третий регистр, переход на который осуществляется передачей комбинации из пяти нулей. Разместить 31 русскую букву

Таблица 2.5

Номер комбинации	Регистры			Кодовая комбинация
	Латинский	Русский	Цифровой	
1	T	Т	5	00001
2	Возврат каретки			00010
3	0	О	9	00011
4	Пробел			00100
5	H	Х	Щ	00101
6	N	Н	,	00110
7	M	М	.	00111
8	Перевод строки			01000
9	L	Л	)	01001
10	R	Р	4 (Ч)	01010
11	G	Г	Ш	01011
12	I	И	8	01100
13	P	П	0	01101
14	C	Ц	:	01110
15	V	Ж	=	01111
16	E	Е	3	10000
17	Z	З	+	10001
18	D	Д	кто там	10010
19	B	Б	?	10011
20	S	С	′ ап	10100
21	Y	Ы	6	10101
22	F	Ф	Э	10110
23	X	Ь	/	10111
24	A	А	—	11000
25	W	В	2	11001
26	J	Й	Ю	11010
27	Цифровой регистр			11011
28	U	У	7	11100
29	Q	Я	1	11101
30	K	К	(	11110
31	Латинский регистр			11111
32	Русский регистр			00000

аналогично латинским не удается, поэтому 5 русских букв помещены в цифровом регистре, 3 из них помещены на свободных позициях цифрового регистра (Ш, Щ и Э), буква «Ю» помещена вместе со звуковым сигналом, а буква «Ч» передается той же комбинацией, что и цифра 4. Заметим, что знаки в табл. 2.5 размещены

таким образом, чтобы соответствующие им кодовые комбинации шли в порядке возрастания номеров этих знаков.

Код № 2 получил повсеместное распространение и до сих пор широко используется в буквопечатающих телеграфных аппаратах.

Развитие вычислительной техники, систем передачи данных, широкое применение средств вычислительной техники для обработки информации сделали необходимой новую стандартизацию первичного кода. Число необходимых служебных и функциональных символов с течением времени значительно возросло, особенно с появлением и развитием систем передачи данных. Появилась необходимость передачи символов различных математических и логических операций, а также ряда служебных знаков (например, «понял», «ждите», «конец адреса» и т. д.).

Отсутствие стандартного кода, который обеспечивал бы возможность получения бо́льшего числа комбинаций, чем в кодах № 1 и 2, создавало большие трудности при передаче данных по каналам связи и их обработке в ЭВМ. Применение различных машинных кодов и кодов для передачи по линиям связи приводило к необходимости использования различной кодопреобразовательной аппаратуры. По мере развития систем передачи данных такое положение становилось экономически все более невыгодным.

Необходимо заметить, что создание нового стандартного кода — очень сложная задача, которая должна учитывать много различных факторов, в том числе очень важный фактор экономики. Введение нового стандартного кода связано с большими затратами на новую аппаратуру, с отказом от ранее существовавшей аппаратуры, с переучиванием обслуживающего персонала и т. п. Новый стандартный код должен быть пригодным как для телеграфии, так и для передачи данных. Он должен быть удобной основой для дальнейшего совершенствования стандартного кода. При разработке нового стандартного кода возникают две основные задачи: 1) выбор множества необходимых кодовых комбинаций, пригодных и для телеграфии, и для передачи данных; 2) оптимальное распределение этого множества в кодовой таблице, т. е. оптимальное составление кодового словаря.

Первая задача связана с увеличением значности

кода, которое вызвано необходимостью расширения кодового словаря для систем передачи данных. Так как в таких системах потребителем цифровой информации является ЭВМ, кодовый словарь должен это учитывать.

Таблица 2.6

b_7 →		0	0	0	0	0	1	1	1	1
b_6 →		0	0	0	1	1	0	0	1	1
b_5 →		0	0	1	0	1	0	1	0	1
$b_7 b_6 b_5$	$b_4 b_3 b_2 b_1$	№ столбца / № строки	0	1	2	3	4	5	6	7
	0 0 0 0	0			Spase	0	@	P	\	p
	0 0 0 1	1	е		!	1	A	Q	a	q
	0 0 1 0	2	ы		"	2	B	R	b	r
	0 0 1 1	3	н		≠&	3	C	S	c	s
	0 1 0 0	4	ь		$	4	D	T	d	t
	0 1 0 1	5	л	ы	%	5	E	U	e	u
	0 1 1 0	6	а	л	&	6	F	V	f	v
	0 1 1 1	7	н	о	/	7	G	W	g	w
	1 0 0 0	8	о	в	(	8	H	X	h	x
	1 0 0 1	9	и	м	)	9	I	Y	i	y
	1 0 1 0	10	ц	и	*	:	J	Z	j	z
	1 0 1 1	11	к	с	+	;	K	([)	k	
	1 1 0 0	12	н		,	<	L		l	
	1 1 0 1	13	у		–	=	M	(])	m	
	1 1 1 0	14	ф		.	>	N	^	n	
	1 1 1 1	15			/	?	O	–	o	DEL

I зона (столбцы 0–1); II зона (столбцы 2–3); III зона (столбцы 4–5); IV зона (столбцы 6–7)

В результате длительной работы (1962 — 1965 гг.) был предложен новый первичный стандартный код для передачи информации по линиям связи и ее обработки в ЭВМ. Это семизначный двоичный код. Предложенный код весьма удобен для стран, где применяется латинский алфавит. Код содержит достаточное число символов для качественной печати телеграфных текстов с латинским алфавитом, для передачи данных и решения наиболее распространенных задач на ЭВМ.

Упрощенная структура нового международного кода для передачи и обработки дискретной информации приведена в табл. 2. 6. Кодовая таблица нового первичного кода условно может быть разделена на четыре зоны. В первой зоне (столбцы 0 и 1) размещены функциональные знаки, которые служат для управления передачей информации по каналам связи (например, «синхронизация», «начало заголовка», «начало текста», «конец текста» и т. п.), для управления печатью («возврат на шаг», «вертикальная табуляция», «горизонтальная табуляция», «возврат каретки» и т. п.), для управления оконечными устройствами («стоп» и другие символы управления), для разделения информации.

Во второй зоне (столбцы 2 и 3) размещены специальные математические символы (+, —, <, >, = и т. п.), знаки пунктуации и цифры от 0 до 9.

В третьей и четвертой зонах (столбцы 4—7) размещены прописные и строчные латинские буквы, а также резервные позиции, которые могут быть использованы национальными ведомствами для своих целей. Знак DEL соответствует операции «забой». Его основным назначением является исключение ошибочных или нежелательных знаков. Как видно из таблицы, стандартный код строится так, что все множество кодовых комбинаций разбивается на два подмножества. Для передачи и обработки чисто цифровой информации три старших элемента кода (b_5, b_6, b_7) отбрасываются и используется только 4-разрядное подможество комбинаций, образуемых символами (b_1, b_2, b_3, b_4). Десятичные цифры (0, 1,..., 9) кодируются комбинациями четырехзначного натурального двоичного кода. Передача любых чисел ведется в двоично-десятичном коде (тетрадами).

В странах с нелатинским алфавитом применение нового стандартного кода связано с определенными трудностями, так как в таблице нового кода нет места для национального алфавита. Поэтому приходится вводить в код еще один регистр, который включал бы национальный алфавит.

С учетом этих обстоятельств в СССР был разработан новый первичный стандартный код. Структура этого кода практически совпадает со структурой Международного стандартного кода. Отличие состоит в том, что разработанный код является двухрегистровым — первый

регистр соответствует латинскому алфавиту, а второй — русскому. Несмотря на недостатки регистровых кодов, которые проявляются в том, что искажение комбинации адреса регистра приводит к неверному приему всех последующих комбинаций вплоть до следующего переключения регистра, разработанный код был утвержден в качестве стандарта СССР и вступил в силу с 1 января 1968 г. Этот код должен применяться как единый первичный код для передачи и обработки телеграфной информации и данных, т. е. информации, предназначенной для обработки в вычислительных машинах или уже обработанной ими.

Введение нового стандартного первичного кода связано с преодолением ряда технических и экономических трудностей (создание новой аппаратуры и замена старой, обучение персонала и т. п.), поэтому внедряется он постепенно. Применяющийся стандартный первичный код № 2 и соответствующая аппаратура не отменяются немедленно и будут использоваться наряду с принятым кодом и новой аппаратурой.

Детальное рассмотрение различных вопросов, связанных с первичными стандартными кодами, содержится в работе [22].

2.7. КРАТКИЕ ИТОГИ ГЛ. 2

1. В этой главе были рассмотрены вопросы, связанные с представлением информации в цифровой форме. К их числу относятся дискретизация непрерывных сообщений по времени и по уровню; выбор параметров цифровых последовательностей, образованных из дискретизированных сообщений; методы получения первичных кодов; оценка погрешностей, возникающих при представлении непрерывных сообщений в цифровой форме, и т. д. Принципиальная важность этих вопросов состоит в том, что от их решения существенно зависят параметры проектируемых систем связи и показатели качества их работы.

2. Рассмотрение показало, что задача дискретизации непрерывных сообщений по времени не имеет однозначного решения. Выбор необходимого интервала дискретизации зависит не только от допустимой погрешности дискретизации, но и от способа восстановления не-

прерывного сообщения по его отсчетам. Для решения этих вопросов необходимо опираться на математическую модель непрерывных сообщений. Наиболее широкое распространение получила модель в виде случайного процесса с ограниченным спектром. Такая модель, несмотря на некоторые принципиальные недостатки, проста и удобна для применения.

3. Для оценки предельных возможностей дискретизации по времени очень важное значение имеет теорема Котельникова. Эта теорема (в рамках исходных допущений) определяет принципиальные возможности дискретизации и указывает пути, позволяющие приближаться к этим возможностям. Однако применение теоремы Котельникова к практическим задачам связано с серьезными трудностями и требует определенной осторожности, так как в реальных условиях допущения, на которых основана теорема, не могут быть строго выполнены.

4. В практических приложениях на выбор интервала дискретизации по времени существенное значение оказывают соображения о простоте и удобстве интерполирующих устройств. Исследования показывают, что линейные и квадратичные интерполяторы позволяют выбирать такие интервалы дискретизации по времени, при которых избыточность отсчетов по сравнению с предельным случаем оказывается не слишком большой.

5. Чтобы преобразовать непрерывное сообщение в дискретное, помимо дискретизации по времени необходима дискретизация по уровню (квантование). Наиболее распространено равномерное квантование по уровню. При таком квантовании среднеквадратичная ошибка определяется числом уровней квантования и практически не зависит от статистики квантуемых сообщений.

6. Для передачи дискретизированных сообщений по линиям связи их преобразуют в цифровую последовательность, т. е. представляют в виде последовательности ряда кодовых комбинаций, являющихся сочетаниями из некоторого конечного числа символов. Из практических соображений весьма удобно основание первичного кода выбирать двоичным. Представление информации в виде цифровой последовательности, формируемой на основе двоичных кодов, в настоящее время наиболее распространено.

7. В качестве элементов двоичных кодов находят применение разнообразные символы, которые формируются

различными способами. Эти способы принято делить на две группы: 1) формирование элементов кода методом возврата к нулевому уровню; 2) формирование элементов кода без возврата к нулевому уровню. Большое разнообразие символов и разные методы их формирования позволяют получать первичные коды в форме, наиболее целесообразной для данных конкретных условий работы и предъявляемых требований.

8. Все первичные коды независимо от формы, в которой они представлены, разделяют на взвешенные и невзвешенные. Наиболее распространенным видом взвешенных кодов является натуральный двоичный код, а невзвешенных — код Грея. Натуральный двоичный код удобен для различных преобразований, ввода в ЭВМ, при декодировании и т. п., а код Грея — для передачи по линиям связи.

9. При развитии различных систем передачи цифровой информации важное значение имеют вопросы стандартизации параметров систем, применяемых кодов, аппаратуры и т. д. В настоящее время такая стандартизация широко применяется при передаче данных. В этой области применяются только стандартные первичные коды.

КОНТРОЛЬНЫЕ ВОПРОСЫ И ЗАДАЧИ

2.1. Какие задачи возникают при дискретизации непрерывных сообщений и в чем их суть?

2.2. Показать, что спектр дискретизированного по времени сообщения определяется выражением

$$A_{д}(\Omega) = \sum_{k=-\infty}^{\infty} u(k\Delta t) \exp(-jk\Delta t\Omega) =$$

$$= \frac{1}{\Delta t} \sum_{k=-\infty}^{\infty} A\left(\Omega - k\frac{2\pi}{\Delta t}\right),$$

где $u(t)$ — непрерывное сообщение, подвергаемое дискретизации;

$A(\Omega)$ — спектр этого сообщения; Δt — интервал дискретизации.

2.3. Указать основные особенности модели непрерывных сообщений с ограниченным спектром. Какие принципиальные недостатки свойственны такой модели?

2.4. Указать основные особенности предельной дискретизации по Котельникову. Какие причины препятствуют ее практическому применению?

2.5. Показать, что если энергетический спектр сообщения определяется выражением

$$G_x(\Omega) = G_0/[1 + (\Omega\tau)^2],$$

то частота дискретизации, найденная в соответствии с теоремой Котельникова, равна

$$F_д = \frac{1}{\Delta t} = \frac{1}{\pi\tau} \operatorname{tg}\left[\frac{\pi}{2}(1 - \delta^2_д)\right] \approx \frac{2}{\pi^2\tau\delta^2_д}.$$

2.6. Решить предыдущую задачу для случая, когда энергетический спектр сообщения определяется выражением

$$G_x(\Omega) = G_0 \exp(-\alpha^2\Omega^2).$$

2.7. Получить формулы, определяющие отношение частот дискретизации по Котельникову для сообщений со спектрами, рассматриваемыми в задачах 2.5 и 2.6. Задачу решить для условий: а) средняя мощность сообщений одинакова; б) ширина энергетических спектров на уровне 0,5 одинакова.

2.8. Основываясь на выражении (2.17), доказать справедливость равенства

$$\int_0^{T_c} x^2(t)dt = \frac{1}{2F_в} \sum_{k=1}^{m} x^2(k\Delta t).$$

2.9. Получить выражения для коэффициента k_x (см. § 2.3), если распределение мгновенных значений квантуемых сообщений подчиняется закону: а) арксинуса; б) треугольника (закон Симпсона); в) Гаусса; г) Релея.

Примечание. Два последних закона «усечь» на уровне «трех сигма».

2.10. Какими преимуществами обладает троичная система счисления по сравнению с другими системами счисления? Почему для передачи цифровой информации в основном применяется двоичная система счисления?

2.11. Какой код называют первичным? Какой максимальной избыточностью может обладать такой код?

2.12. Дать сравнительную характеристику натурального двоичного кода (НДК) и кода Грея.

2.13. Предложить структуру кодопреобразователей: 1) НДК в код Грея; 2) код Грея в НДК.

СПИСОК ЛИТЕРАТУРЫ

1. **Свенсон А. Н., Смердов А. А.** Системы передачи информации со статистическим усреднением. Киев, «Наукова думка», 1967.

2. **Ольховский Ю. Б., Новосёлов О. Н., Мановцев А. П.** Сжатие данных при телеизмерениях. М., «Сов. радио», 1971.

3. **Сокращение** избыточности. — ТИИЭР, Тематический выпуск. № 3. М., «Мир», 1967.

4. **Стейн С., Джонс Дж.** Принципы современной теории связи и их применение к передаче дискретных сообщений. М., «Связь», 1971.

5. **Левин Б. Р.** Теоретические основы статистической радиотехники. Книга перва. М., «Сов. радио», 1966.

6. **Харкевич А. А.** О теореме Котельникова (обзор некоторых работ). — «Радиотехника», 1958, № 8.

7. **Железнов Н. А.** Некоторые вопросы теории информационных электрических систем. Л., ЛКВВИА им. А. Ф. Можайского, 1960.

8. **Хургин Я. И., Яковлев В. П.** Финитные функции в физике и технике. М., «Наука», 1971.

9. **Трахтман А. М.** Введение в обобщенную спектральную теорию сигналов. М., «Сов. радио», 1972.

10. **Теория** информации и ее приложения. (Сб. переводов с англ. Под ред. А. А. Харкевича.) М., ГИФМЛ, 1959.

11. **Шеннон К.** Работы по теории информации и кибернетике. М., ИЛ, 1963.

12. **Борисов Ю. П., Пенин П. И.** Основы многоканальной передачи информации. М., «Связь», 1967.

13. **Броун В. М.** Анализ линейных инвариантных во времени систем. М., Машиностроение», 1966.

14. **Мановцев А. П.** Основы теории радиотелеметрии. М., «Энергия», 1973.

15. **Пенин П. И.** Оценка избыточности при некоторых видах равномерной дискретизации непрерывных сообщений. Доклады научно-технической конференции МЭИ. Подсекция системы связи. М., МЭИ, 1969.

16. **Маркюс Ж.** Дискретизация и квантование. Пер. с фанц. М., «Энергия», 1969.

17. **Солодов А. В.** Теория информации и ее применение к задачам автоматического управления и контроля. М., «Наука», 1967.

18. **Кэтермоул К. В.** Принципы импульсно-кодовой модуляции. Пер. с англ. М., «Связь», 1974.

19. **Дупраз Ж.** Спектры телеметрических КИМ-сигналов. Пер. с франц. — «Радиолокация, телевидение и радиосвязь», 1968, № 12.

20. **Справочник** по телеметрии. Пер. с англ. Под ред. Р. Т. Сафарова. М., «Машиностроение», 1971.

21. **Юргенсон Р. И.** Помехоустойчивость цифровых систем передачи телемеханической информации. Л., «Энергия», 1971.

22. **Супрун Б. А.** Первичные коды. М., «Связь», 1970.

23. **Шляпоберский В. И.** Основы техники передачи дискретных сообщений. М., «Связь», 1973.

24. **Субье-Ками.** Двоичная техника и обработка информации. М., «Мир», 1964.

Глава 3

Оптимальный прием двоичных сигналов в каналах с постоянными параметрами

3.1. ВВЕДЕНИЕ

В гл. 2 было выяснено, что любые сообщения дискретного или непрерывного характера можно преобразовать в цифровую форму, т. е. представить в виде последовательности соответствующих кодовых комбинаций некоторого первичного кода. Следовательно, термин «цифровая форма» совсем не означает, что все передаваемые сообщения являются числами, а лишь указывает на то, что эти сообщения представлены дискретными последовательностями, элементы которых могут принимать лишь некоторое конечное множество значений.

В этой главе будем рассматривать последовательности, элементы которых могут принимать только два возможных значения. Если имеются достаточно протяженные во времени двоичные последовательности, можно показать, что для любого кода в среднем число элементов одного и другого вида (например, нулей и единиц) в таких последовательностях одинаково или почти одинаково. Поэтому цифровую информацию, поступающую в канал связи, можно рассматривать как последовательность нулей и единиц, вырабатываемую дискретным источником, статистическая схема которого имеет вид

$$x = \begin{pmatrix} x_1 & x_2 \\ p_1 & p_2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0{,}5 & 0{,}5 \end{pmatrix}. \tag{3.1}$$

Здесь p_1 и p_2 — вероятности появления нуля и единицы в двоичной последовательности.

Это простейшая статистическая схема источника, в которой не учитываются статистические особенности различных сочетаний символов, например, вида 00; 01; 10; 111 и т. п., которые могут быть описаны соответствующими матрицами условных вероятностей.

Схема (3.1) достаточна для исследования различных вопросов передачи и приема двоичных цифровых по-

следовательностей по каналам с постоянными параметрами, в которых единственным видом помех является белый гауссовский шум. При таком подходе при передаче информации учитывается только влияние выбранного вида сигналов и действующих в канале помех и не учитываются особенности статистического и помехоустойчивого кодирования. Иными словами, в этой главе рассматриваются вопросы передачи отдельных элементов двоичных последовательностей по каналу связи без

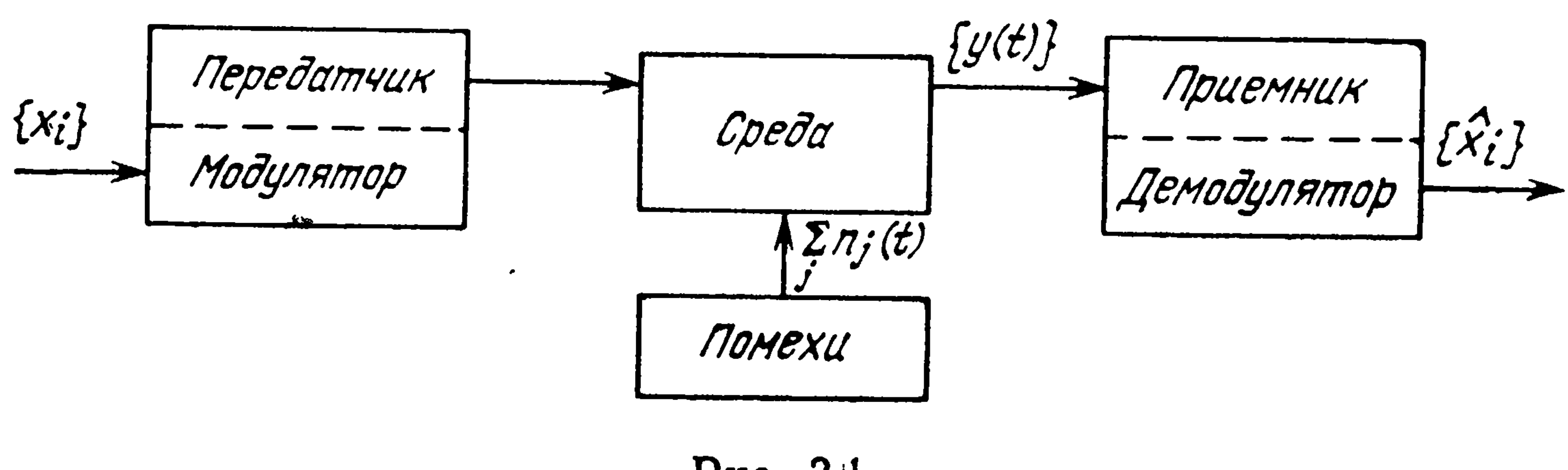


Рис. 3.1.

кодирования, который представляет собой неизменяемую часть цифровой линии связи и содержит только передатчик (включая модулятор), среду распространения и приемник (включая демодулятор).

Структурная схема цифрового канала связи без кодирования показана на рис. 3.1. Задача заключается в том, чтобы передать входную двоичную последовательность $\{x_i\}$ на выход канала. Чтобы количественно учесть влияние помех и других факторов, которые могут вызывать отличие принятой последовательности $\{\hat{x}_i\}$ от переданной, необходимо ввести критерий оценки качества принятой информации. При передаче информации в цифровой форме за такой критерий обычно принимают вероятность ошибки приема одного элемента цифровой последовательности*.

Так как каждый элемент двоичной последовательности может принимать лишь одно из двух возможных значений, то для передачи последовательности достаточно иметь два различных сигнала $s_1(t)$ и $s_2(t)$, длительность каждого из которых равна длительности элемента

---

* Качество принятой цифровой информации оценивают вероятностью ошибки, а не вероятностью правильного приема потому, что последняя обычно очень близка к единице и поэтому удобнее брать разность между единицей и вероятностью правильного приема ($p_{\text{ош}}=1-p_{\text{пр}}$).

τ_0. Передающая часть системы как бы заменяет цифровую последовательность $\{x_i\}$ на соответствующую последовательность сигналов $\{s_i(t)\}$. Эта замена осуществляется в результате дискретного изменения одного из параметров синусоидального сигнала передатчика в соответствии с передаваемой цифровой последовательностью. Такое изменение параметра сигнала называется манипуляцией. Манипулировать можно амплитуду, частоту или фазу несущего колебания передатчика. В простейшем случае, когда изменяемый параметр сигнала передатчика может принимать только два состояния, имеем наиболее распространенный вид манипуляции — бинарную манипуляцию.

Помимо полезных информационных изменений, сигнал, приходящий в точку приема, может содержать некоторые паразитные изменения параметров из-за неидеальных условий распространения и других причин. Изменения параметров могут быть вызваны неоднородностями и случайностями характеристик среды распространения; изрезанностью диаграмм направленности передающей и приемной антенн и их взаимным перемещением в пространстве; влиянием эффекта Допплера; нестабильностью характеристик аппаратуры.

В общем случае на вход приемного устройства поступает смесь искаженного сигнала $\hat{s}_i(t)$ и различных помех $n_j(t)$, т. е.

$$y(t) = \hat{s}_i(t) + \sum_j n_j(t), \quad 0 \leqslant t \leqslant \tau_0. \tag{3.2}$$

Колебание $y(t)$ представляет собой одну из возможных реализаций некоторого случайного процесса, рассматриваемого на интервале длительностью τ_0.

Какую задачу должен решить приемник? Анализируя принятую конкретную реализацию $y(t)$ в течение интервала времени $0 \leqslant t \leqslant \tau_0$, приемник должен установить, какой из возможных сигналов $s_i(t)$ ($s_1(t)$ или $s_2(t)$) присутствует на его входе, и в соответствии с этим принять решение, какой элемент цифровой последовательности x_i (x_1 или x_2) был передан. Эта классическая задача теории связи — задача различения двух сигналов. Если один из сигналов тождественно равен нулю, имеем частный случай задачи различения — задачу обнаружения сигнала в интервале времени $0 \leqslant t \leqslant \tau_0$ на фоне помех.

Различение сигналов — задача, характерная для систем передачи цифровой информации, дискретных систем телеуправления и т. п.

Для решения этой задачи необходимо располагать рядом исходных данных о применяемых сигналах, свойствах среды распространения и помехах, которые действуют в канале. Эти предварительно известные сведения называются априорными. Априорные сведения можно получить, обрабатывая различные экспериментальные данные, относящиеся к передаче информации по интересующему нас каналу связи. Результаты экспериментального исследования далее кладутся в основу теоретического рассмотрения уже как заранее известные. Заранее известные сведения, представляющие собой априорную информацию об условиях передачи, могут быть полными, неполными или отсутствовать вовсе.

Априорная информация является полной, если известны статистические характеристики всех случайных процессов, поступающих на вход приемника. Практически располагать полной и достоверной априорной информацией почти никогда не удается. Поэтому задачу различения сигналов приходится решать, вводя некоторую идеализацию исходных условий, при которой сохраняются наиболее важные особенности рассматриваемого канала связи. В частности, в ряде каналов связи среда распространения может в первом приближении считаться идеальной, т. е. неискажающей, а ее параметры постоянными. В этом случае все характеристики и свойства сигнала на входе приемника остаются такими же, как и на выходе передатчика *. Такие сигналы называются сигналами, известными точно. В этих сигналах все параметры, за исключением параметра, содержащего полезную информацию, известны на приемной стороне системы связи точно.

Если характеристики среды распространения меняются во времени случайно, параметры сигнала на входе приемника также становятся случайными. Такие сигналы называются сигналами со случайными параметрами (флюктуирующими сигналами).

Строго говоря, законы распределения плотности вероятностей случайных параметров сигналов никогда

---

* Изменяется только уровень сигнала из-за ослабления при распространении, что эквивалентно умножению переданного сигнала на некоторый постоянный коэффициент.

не известны точно. Поэтому при рассмотрении флюктуирующих сигналов обычно задаются заранее некоторым классом распределений, приближенно соответствующих фактическому положению, т. е. прибегают к некоторой идеализации исходных данных. Качество различения сигналов зависит не только от вида сигналов и их характеристик, но и от ряда других факторов, в том числе от вида помех, действующих в канале и их статистических характеристик; искажений сигнала в тракте приемного устройства; точности знания начала и конца временного интервала, на котором осуществляется различение сигналов.

Из большого количества различных помех, которые могут действовать в канале связи, принципиально неустранимы внутренние флюктуационные шумы приемника. Именно эти помехи ограничивают предельные возможности передачи информации по каналам связи. Идеализацией таких помех является белый гауссовский шум со спектральной плотностью N_0.

Искажения сигналов в приемном устройстве проявляются в виде наложения остаточных переходных процессов от предыдущих посылок на последующие и тем самым затрудняют различение этих посылок. Иными словами, эти искажения приводят к некоторой деманипуляции принятых сигналов. Как известно, для уменьшения таких искажений приходится вводить защитные интервалы между символами или расширять полосу пропускания приемника. Естественно, что такие меры будут приводить к ухудшению некоторых показателей качества работы системы связи, в частности, уменьшать скорость передачи или увеличивать затраты полосы на передачу единицы информации. Однако нас в данном случае интересует только тот факт, что при различении сигналов влияние искажений сигналов принципиально может быть сведено к сколь угодно малому значению.

Для различения сигналов в приемнике необходимо иметь возможность точно (или с допустимой погрешностью) устанавливать начало и конец интервала анализа каждой реализации $y(t)$, поступающей на вход приемника. Такая возможность обеспечивается устройством символьной синхронизации, которое позволяет устанавливать начало и конец каждой принимаемой сигнальной посылки.

Проведенное выше рассмотрение позволяет сформулировать основные допущения, которые принимаются при решении задачи различения сигналов в классической постановке:

1) помеха, действующая в канале, представляет собой белый гауссовский шум с односторонней спектральной плотностью N_0;

2) приемное устройство считается неискажающим, т. е. при отсутствии помех информация принимается безошибочно;

3) система синхронизации в приемном устройстве работает идеально, т. е. позволяет точно устанавливать начало и окончание любой посылки сигнала в принятой последовательности.

При таких допущениях необходимо располагать только априорными сведениями о статистических характеристиках принимаемых сигналов.

В этой главе рассматривается различение сигналов, известных точно. Иными словами, рассматривается различение сигналов в канале с постоянными параметрами и аддитивной помехой типа гауссовского белого шума. Такие каналы называют гауссовскими. Они являются достаточно хорошей идеализацией ряда реальных каналов передачи цифровой информации, в частности, космических каналов связи, радиорелейных и кабельных каналов и некоторых других.

Впервые задача различения бинарных сигналов была решена В. А. Котельниковым в 1946 г. Его работа [1] сыграла большую стимулирующую роль в развитии общей теории связи и вызвала появление огромного количества исследований, многие из которых существенно продвинули вперед решение проблемы оптимального приема дискретных сигналов. Даже краткий обзор наиболее крупных достижений в этой области потребовал бы значительного объема и изложения ряда дополнительных сложных вопросов. Поэтому целью данной главы является детальное рассмотрение задачи оптимального различения двоичных сигналов в классической постановке. Помимо целей методологического характера, такое рассмотрение очень важно и для понимания многих более сложных задач оптимального приема, при решении которых делаются попытки снять ряд ограничений, характерных для классической постановки.

3.2. ОПТИМАЛЬНОЕ РАЗЛИЧЕНИЕ ДИСКРЕТНЫХ СИГНАЛОВ МЕТОДОМ ПРОВЕРКИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Основным методом решения задачи различения сигналов является метод проверки статистических гипотез [2—7]. Первоначально этот метод был разработан в математической статистике для испытания различных гипотез, а затем успешно применен при решении радиолокационных задач и задач передачи дискретной информации. Расширение круга теоретических и прикладных задач статистического характера в автоматике, радиотехнике и других областях, а также быстрое развитие математической статистики привело к тому, что в 50-х годах нашего столетия сложился очень важный раздел современной теории связи — теория статистических решений. Эта теория объединила и развила многие методы статистики, теории игр, а также методы, разработанные в прикладных областях знаний. Поэтому метод проверки статистических гипотез является в настоящее время одним из частных случаев современной теории статистических решений. Он очень удобен для решения задач различения дискретных сигналов. Рассмотрим суть метода проверки гипотез применительно к различению сигналов.

При решении задачи различения сигналов ответ должен быть получен на основе априорных сведений о статистике сигналов и анализа входного колебания $y(t)$ на интервале $0 \leqslant t \leqslant \tau_0$. Для сигналов, известных точно, априорные сведения о статистике сигналов задаются вероятностями появления того или иного сигнала на входе приемника, или что то же самое, вероятностями передачи того или иного символа. Это может быть записано в виде статистической схемы

$$\begin{pmatrix} x_1 & x_2 \\ p(x_1) & p(x_2) \end{pmatrix} = \begin{pmatrix} s_1(t) & s_2(t) \\ p(s_1) & p(s_2) \end{pmatrix}, \qquad (3.3)$$

где x_1 и x_2 — символы алфавита, с помощью которого формируются двоичные цифровые последовательности; $s_1(t)$ и $s_2(t)$ — сигналы, с помощью которых по каналу связи передаются символы x_1 и x_2; $p(x_1)$ и $p(x_2)$ — вероятности передачи символов x_1 и x_2; $p(s_1)$ и $p(s_2)$ — вероятности появления сигналов $s_1(t)$ и $s_2(t)$ на входе приемника.

Естественно, что

$$p(x_1)=p(s_1),\ p(x_2)=p(s_2),\ p(s_1)+p(s_2)=1. \quad (3.4)$$

Для канала с постоянными параметрами и аддитивной помехой $n(t)$ колебание на входе приемника имеет вид

$$y(t)=s_i(t)+n(t), \qquad 0\leqslant t\leqslant\tau_0, \quad (3.5)$$

где $s_i(t)$ может принять либо значение $s_1(t)$ с вероятностью $p(s_1)$, либо значение $s_2(t)$ с вероятностью $p(s_2)$.

Задача заключается в том, чтобы, анализируя колебание $y(t)$, получить как можно бóльшую информацию о том, какой входной сигнал $s_i(t)$ присутствует в этом колебании: $s_1(t)$ или $s_2(t)$. Очень важно уяснить, что информация о сигнале $s_i(t)$, которая содержится в колебании $y(t)$, не может быть увеличена никакими операциями над $y(t)$. Чтобы мы ни делали с колебанием $y(t)$, количество информации о сигнале $s_i(t)$ от этого увеличиться не может, оно может только уменьшиться. Следовательно, нужно так анализировать колебание $y(t)$, чтобы наилучшим образом сохранить имеющуюся полезную информацию о сигнале $s_i(t)$, исключая из колебания $y(t)$ возможно бóльшую часть ненужной (ложной) информации, обусловленной помехой $n(t)$. Выполнив подобный анализ, приемное устройство должно решить, какой сигнал был принят (s_1 или s_2), а следовательно, какой элемент цифровой последовательности был передан (x_1 или x_2). Это статистическая задача — истинной может быть только одна из двух возможных гипотез: 1) гипотеза H_1 — принят сигнал s_1; 2) гипотеза H_2 — принят сигнал s_2.

Выбор той или иной гипотезы может быть правильным, а может быть ошибочным. Очевидно, что при передаче сигналов s_1 и s_2 возможны четыре варианта выбора гипотез H_1 и H_2. Эти варианты представлены в табл. 3.1.

Выбор гипотезы должен быть основан на некотором заранее установленном правиле. Это правило не может быть получено из каких-либо предпосылок, оно принимается как постулат, основанный на соображениях здравого смысла.

Как же сформулировать это правило? Колебание $y(t)$ на входе приемника, кроме шума $n(t)$, содержит либо сигнал s_1, либо сигнал s_2. Обозначим вероятность того, что в поступившем на вход приемника колебании

$y(t)$ содержится сигнал s_1 через $p(s_1|y)$, а сигнал s_2 — через $p(s_2|y)$. Это условные вероятности. Пока колебание $y(t)$ не проанализировано приемником, эти вероятности неизвестны. Предположим, что анализ колебания $y(t)$ проведен и эти вероятности стали известны, т. е. кроме имеющихся априорных сведений появились дополни-

Таблица 3.1

Передан сигнал	Выбрана гипотеза о сигнале	
	$s_1(t)$	$s_2(t)$
$s_1(t)$	Гипотеза H_1 верна	Гипотеза H_2 ошибочна
$s_2(t)$	Гипотеза H_1 ошибочна	Гипотеза H_2 верна

тельные сведения о сигналах на входе приемника. Эти новые, появившиеся в результате анализа сведения, называются апостериорными*. Итак, допустим, что нам известны апостериорные вероятности $p(s_1|y)$ и $p(s_2|y)$. Тогда разумным правилом выбора гипотез является следующее: после того, как проведен анализ поступившего на вход колебания $y(t)$, следует выбрать гипотезу H_1, если вероятность $p(s_1|y)$ превышает вероятность $p(s_2|y)$, и наоборот, принять гипотезу H_2, если $p(s_1|y)$ меньше $p(s_2|y)$.

Следовательно, правило таково:

$$\begin{aligned} p(s_1|y) &> p(s_2|y) \text{ принимается гипотеза } H_1, \\ p(s_1|y) &< p(s_2|y) \text{ принимается гипотеза } H_2. \end{aligned} \tag{3.6}$$

Если $p(s_1|y)=p(s_2|y)$, может быть принята любая из гипотез. В таком случае заранее оговаривают, какую из них принимать.

Правило можно записать в иной, более удобной форме. Введем величину

$$\Lambda_{12}=p(s_1|y)/p(s_2|y). \tag{3.7}$$

* От латинского термина «a posteriori» — из последующего, из опыта, на основании опыта.

Тогда правило принимает вид

$$\Lambda_{12} \gtrless 1 \quad \begin{matrix} \text{принимается гипотеза } H_1, \\ \text{принимается гипотеза } H_2. \end{matrix} \qquad (3.8)$$

Величина Λ_{12} характеризует правдоподобность той или иной гипотезы. Например, при $\Lambda_{12}>1$ полагается, что гипотеза H_1 правдоподобнее гипотезы H_2. Поэтому величину Λ_{12} называют отношением правдоподобия.

Правило выбора гипотез, основанное на том, что выбирается та гипотеза, которой соответствует большая апостериорная вероятность, называется байесовским правилом, или байесовским решением [6].

В литературе такой подход к выбору гипотез часто называют принципом (или критерием) максимума апостериорной вероятности [4, 5, 7], поскольку предпочтение отдается гипотезе, которой соответствует максимальное значение апостериорной вероятности. Байесовское правило решения автоматически обеспечивает минимизацию вероятности ошибки при выборе гипотез и, следовательно, является наиболее разумным правилом при оценке качества передачи информации в системах связи.

Действительно, пусть имеем условие

$$p(s_1|y)>p(s_2|y), \qquad (3.9)$$

в соответствии с которым принято решение, что был передан сигнал s_1. Так как

$$p(s_1|y)+p(s_2|y)=1, \qquad (3.10)$$

величина

$$p(s_2|y)=1-p(s_1|y) \qquad (3.11)$$

характеризует вероятность ошибки такого решения. Если при условии (3.9) принять решение о том, что был передан сигнал s_2, вероятность ошибки такого решения равна

$$p(s_1|y)=1-p(s_2|y). \qquad (3.12)$$

С учетом неравенства (3.9) эта вероятность ошибки будет больше, чем в первом случае, и, следовательно, надо выбрать первую гипотезу. Таким образом, байесовское правило решения обеспечивает минимальную вероятность ошибки при выборе гипотез и с этой точки зрения является оптимальным правилом.

Для того чтобы применить изложенное правило решения к задаче различения двух сигналов, необходимо знать условные вероятности $p(s_1|y)$ и $p(s_2|y)$. Поэтому следующим этапом решения задачи является определение этих вероятностей. Для этого можно воспользоваться одной из модификаций формулы Байеса (см., например, [7]), которая в нашем случае может быть представлена в виде

$$p(s_i|y)=p(s_i)W(y|s_i)/W(y), \tag{3.13}$$

где s_i соответствует сигналу s_1 или s_2; $W(y|s_i)$ можно трактовать как многомерную функцию плотности распределения случайного колебания $y(t)$ при данном фиксированном значении сигнала s_i на входе приемника (условная плотность распределения колебания $y(t)$); $W(y)$ — многомерная безусловная функция распределения колебания $y(t)$. Заметим, что здесь и ниже колебание y необходимо рассматривать как вектор, определяемый своими координатами в m-мерном пространстве: $\mathbf{y}=(y_1, y_2, \ldots, y_m)$, что эквивалентно представлению $y(t)$ в виде совокупности мгновенных значений, взятых в моменты времени $t_1, t_2, \ldots, t_m$. Такой подход характерен для задач статистического приема сигналов. Подробно эти вопросы изложены в работах [3—7], и мы на них не останавливаемся.

Формулу Байеса часто называют формулой обратной вероятности, а вероятности $p(s_1|y)$ и $p(s_2|y)$ — обратными вероятностями. Этот термин объясняется тем, что формула Байеса дает возможность по законам распределения следствия (колебание $y(t)$) найти вероятность причины (сигнал s_i), которая вызвала данное следствие.

Образуем отношения в соответствии с определением (3.7) и формулой (3.13):

$$\Lambda_{12}=\frac{p(s_1|y)}{p(s_2|y)}=\frac{p(s_1)}{p(s_2)}\,\frac{W(y|s_1)}{W(y|s_2)}, \tag{3.14}$$

$$\Lambda_{21}=\frac{p(s_2|y)}{p(s_1|y)}=\frac{p(s_2)}{p(s_1)}\,\frac{W(y|s_2)}{W(y|s_1)}. \tag{3.15}$$

Нетрудно видеть, что всегда должно выполняться условие

$$\Lambda_{12}\Lambda_{21}=1. \tag{3.16}$$

Чтобы конкретизировать эти выражения, необходимо знать функции $W(y|s_1)$ и $W(y|s_2)$. Рассмотрим сначала функцию $W(y|s_1)$. Так как сигнал s_1 является детерминированным, т. е. все его параметры известны точно, то очевидно, что в данном случае закон распределения колебания $y(t)=s_1(t)+n(t)$ полностью определяется многомерным законом распределения помехи $n(t)$, которая присутствует в колебании $y(t)$. Следовательно,

$$W(y|s_1)=W_m(n)=W_m(y-s_1). \qquad (3.17)$$

Аналогично можно записать и для сигнала s_2

$$W(y|s_2)=W_m(n)=W_m(y-s_2). \qquad (3.18)$$

В этих выражениях $W_m(n)$ — многомерное распределение нормального белого шума, рассматриваемого в некоторой полосе частот Δf.

Как известно, такое распределение может быть записано в виде [3, 4]

$$W_m(n)=\prod_{k=1}^{m} W_1(n)=\frac{1}{(2\pi\sigma^2_{ш})^{m/2}}\exp\left(-\frac{\sum\limits_{k=1}^{m} n^2_k}{2\sigma^2_{ш}}\right), \qquad (3.19)$$

где m — размерность многомерного пространства шума; $W_1(n)$ — одномерная плотность распределения шума; $\sigma^2_{ш}$ — средний квадрат шума (мощность шума в полосе Δf).

Выражение (3.19) представляет собой определенную идеализацию, так как оно предполагает замену непрерывных реализаций белого шума набором случайных независимых значений этого шума, взятых в m точках интервала анализа $0\leqslant t\leqslant\tau_0$. Эти значения берутся в соответствии с теоремой Котельникова, и их число равно

$$m=2\Delta f\tau_0. \qquad (3.20)$$

Чем больше величина m, тем точнее выражение (3.19) определяет многомерное распределение белого шума. С физической точки зрения требуемое условие ($m\gg 1$) означает, что искажения сигнала до начала анализа колебания $y(t)$ на интервале времени $0\leqslant t\leqslant\tau_0$ должны полностью отсутствовать. Естественно, что искажения сигнала до начала анализа тем меньше, чем шире полоса частот Δf, в которой рассматривается колебание $y(t)$.

Можно показать, что при достаточно больших значениях m ($m \gg 1$), что соответствует условию, когда шум может считаться белым, справедливо равенство

$$\frac{1}{2\Delta f}\sum_{k=1}^{m} n^2{}_k = \int_0^{\tau_0} n^2(t)\,dt. \qquad (3.21)$$

Тогда выражение (3.19) принимает вид

$$W_m(n) = \frac{1}{(2\pi\sigma^2{}_{\text{ш}})^{m/2}} \exp\left[-\frac{1}{N_0}\int_0^{\tau_0} n^2(t)\,dt\right], \qquad (3.22)$$

где учтено, что $\sigma^2{}_{\text{ш}} = N_0\Delta f$.

Принимая во внимание (3.22), имеем

$$W(y|s_1) = \frac{1}{(2\pi\sigma^2{}_{\text{ш}})^{m/2}} \exp\left\{-\frac{1}{N_0}\int_0^{\tau_0}[y(t) - s_1(t)]^2\,dt\right\}, \qquad (3.23)$$

$$W(y|s_2) = \frac{1}{(2\pi\sigma^2{}_{\text{ш}})^{m/2}} \exp\left\{-\frac{1}{N_0}\int_0^{\tau_0}[y(t) - s_2(t)]^2\,dt\right\}. \qquad (3.24)$$

Используя эти выражения, образуем отношение правдоподобия

$$\Lambda_{12} = \frac{p(s_1)}{p(s_2)} \exp\left\{\frac{1}{N_0}\int_0^{\tau_0}[y(t) - s_2(t)]^2\,dt - \frac{1}{N_0}\int_0^{\tau_0}[y(t) - s_1(t)]^2\,dt\right\}. \qquad (3.25)$$

Полученное выражение для Λ_{12} принципиально уже решает задачу оптимального различения двух сигналов. Действительно, в этом выражении функционально связаны все известные и неизвестные величины. Поэтому, зная Λ_{12} и поступая в соответствии с байесовским правилом, можно решить, какой из двух возможных сигналов (s_1 или s_2) был передан, а следовательно, какой из элементов сообщения (x_1 или x_2) необходимо воспроизвести на выходе канала. Сказанное можно условно записать так:

$$\Lambda_{12} \gtrless 1 \quad \begin{array}{l}\text{принят сигнал } s_1(t),\\ \text{принят сигнал } s_2(t).\end{array} \qquad (3.26)$$

Таким образом, (3.26) с учетом (3.25) можно рассматривать как алгоритм оптимальной процедуры различения двух сигналов на фоне белого гауссовского шума *.

3.3. СТРУКТУРА ОПТИМАЛЬНЫХ ПРИЕМНИКОВ РАЗЛИЧЕНИЯ ДВУХ СИГНАЛОВ НА ФОНЕ БЕЛОГО ГАУССОВСКОГО ШУМА

Результаты, полученные в предыдущем параграфе, дают возможность установить структуру оптимального приемника различения двух сигналов на фоне белого гауссовского шума. Однако полученный алгоритм довольно сложен и неудобен для применения. Для упрощения этого алгоритма изменим форму байесовского правила (3.26), логарифмируя обе части неравенства. Тогда правило примет следующую форму:

$$\ln \Lambda_{12} \gtrless 0 \quad \begin{matrix} \text{принят сигнал } s_1, \\ \text{принят сигнал } s_2. \end{matrix} \tag{3.27}$$

Такое преобразование не меняет сути правила: решение, определяемое неравенствами (3.27), всегда будет таким же, как и решение на основании неравенств (3.26), поскольку логарифм величины Λ_{12} является монотонно изменяющейся функцией этой величины и, следовательно, сохраняет все особенности ее изменения. Взяв логарифм отношения правдоподобия (3.25) и поступив в соответствии с правилом (3.27), можно получить алгоритм оптимального различения двух сигналов в виде

$$\frac{1}{N_0}\int_0^{\tau_0}[y(t)-s_2(t)]^2\,dt-\frac{1}{N_0}\int_0^{\tau_0}[y(t)-s_1(t)]^2\,dt \gtrless$$

$$\gtrless \ln\frac{p(s_2)}{p(s_1)} \quad \begin{matrix} \text{принят сигнал } s=s_1, \\ \text{принят сигнал } s=s_2. \end{matrix} \tag{3.28}$$

---

* Алгоритм (алгорифм) — система правил, следуя которым можно получить решение поставленной задачи путем последовательного выполнения определенных вычислительных операций. Истоки происхождения термина ведут начало от имени арабского математика IX в. аль-Хорезми.

Если учесть, что при передаче информации в двоичной цифровой форме почти всегда имеет место условие

$$p(s_1) \approx p(s_2) = 0{,}5, \tag{3.29}$$

т. е. передаваемые сигналы равновероятны, то выражение (3.28) упрощается:

$$\int_0^{\tau_0} [y(t) - s_2(t)]^2\, dt \gtrless \int_0^{\tau_0} [y(t) - s_1(t)]^2\, dt \quad \begin{matrix} \text{принят сигнал } s = s_1, \\ \text{принят сигнал } s = s_2. \end{matrix} \tag{3.30}$$

Это выражение имеет ясный физический смысл: наиболее вероятным переданным сигналом является тот сигнал, который меньше отличается (в среднеквадра-

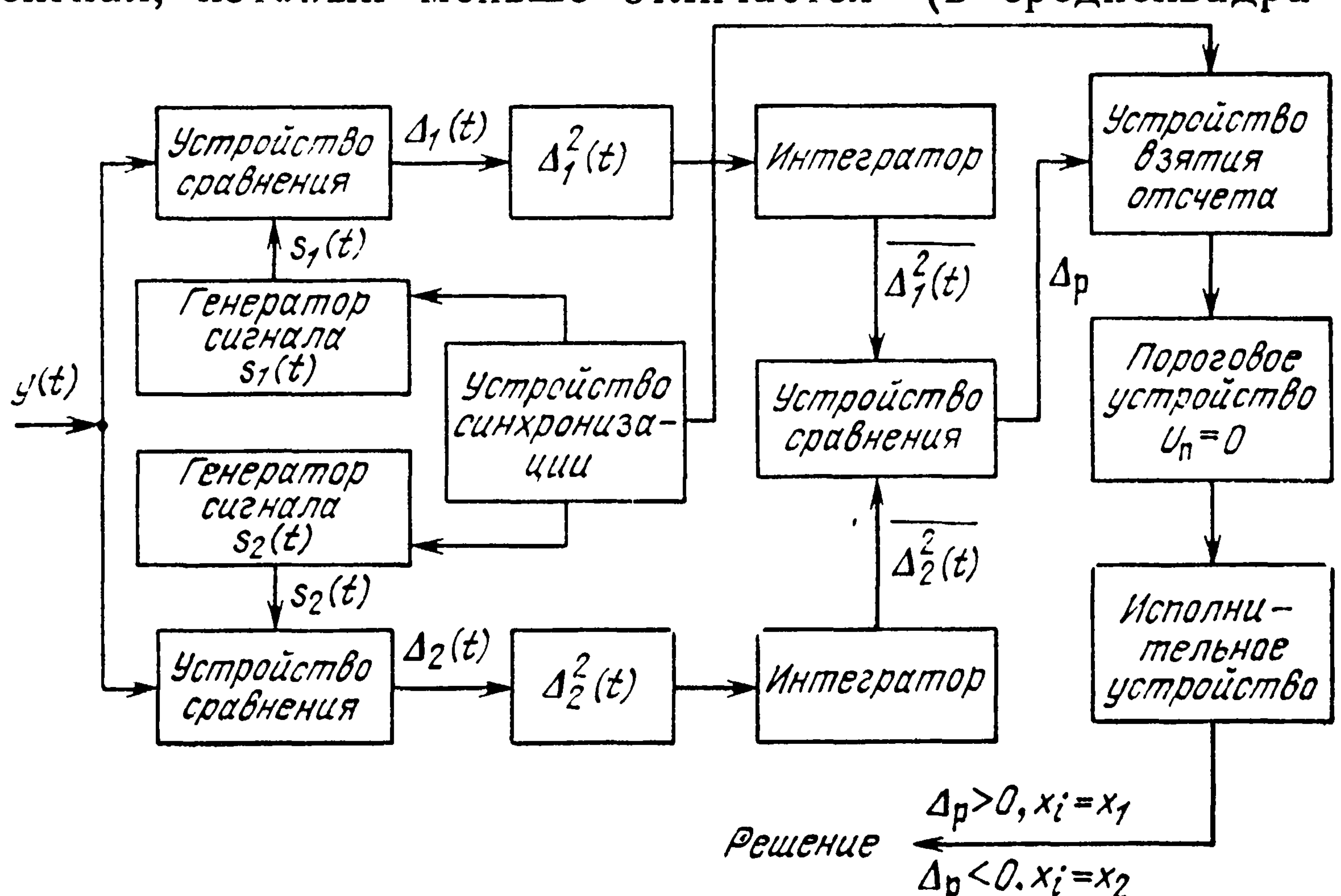


Рис. 3.2.

тичном смысле) от входного колебания $y(t)$. В соответствии с этим алгоритмом можно построить оптимальный приемник различения двух сигналов, т. е. представить найденный алгоритм в виде некоторой структурной схемы.

Структура оптимального приемника в соответствии с (3.30) показана на рис. 3.2. Приемник имеет две ветви, в каждой из которой вычисляется среднеквадратичная величина отличия принятого колебания $y(t)$ от известного сигнала: в первой ветви от сигнала $s_1(t)$, во

второй — от $s_2(t)$. Решение принимается по той ветви, где это отличие меньше. Рассмотренная структура не единственна для оптимального приемника.

Введем обозначения

$$\zeta(s_i) = \int_0^{\tau_0} y(t)\, s_i(t)\, dt, \quad i = 1,\ 2; \tag{3.31}$$

$$E_i = \int_0^{\tau_0} s_i^2(t)\, dt, \quad i = 1,\ 2. \tag{3.32}$$

Величина ζ характеризует взаимную корреляцию между $y(t)$ и $s_i(t)$ и называется корреляционным интегралом. Величина E_i определяет удельную энергию (энергию на активном сопротивлении 1 Ом) сигнала $s_i(t)$. С учетом этих обозначений выражение (3.30) можно преобразовать к виду

$$\zeta(s_1) - \zeta(s_2) \gtrless \frac{E_1 - E_2}{2} \quad \begin{matrix} \text{принят сигнал } s = s_1, \\ \text{принят сигнал } s = s_2. \end{matrix} \tag{3.33}$$

Далее рассмотрим два важных частных случая.

3.3.1. Оптимальные приемники сигналов с пассивной паузой

Сигналом с пассивной паузой называют такой сигнал, который излучается только при передаче информационных символов 1. Передаче информационных символов 0 соответствуют паузы между посылками сигнала. Следовательно,

$$E_1 = E \ (s_1(t) \neq 0); \quad E_2 = 0 \ (s_2(t) \equiv 0). \tag{3.34}$$

Правило оптимального различения (3.33) в этом случае принимает вид

$$\zeta(s_1) \gtrless \frac{E}{2} \quad \begin{matrix} \text{принят сигнал } s = s_1, \\ \text{принят сигнал } s = s_2 \end{matrix} \tag{3.35}$$

или с учетом (3.31)

$$\zeta(s_1) = \int_0^{\tau_0} y(t)\, s_1(t)\, dt \gtrless \frac{E}{2} \quad \begin{matrix} \text{принят сигнал } s = s_1, \\ \text{принят сигнал } s = s_2. \end{matrix} \tag{3.36}$$

Это выражение представляет собой алгоритм работы оптимального приемника сигналов с пассивной паузой.

В соответствии с этим правилом приемник должен вычислять значение корреляционного интеграла и сравнивать его с некоторой фиксированной величиной (порогом).

Подобная операция может быть выполнена с помощью схемы, показанной на рис. 3.3. Схема представляет собой оптимальный корреляционный приемник, состоящий из коррелятора и решающего устройства.

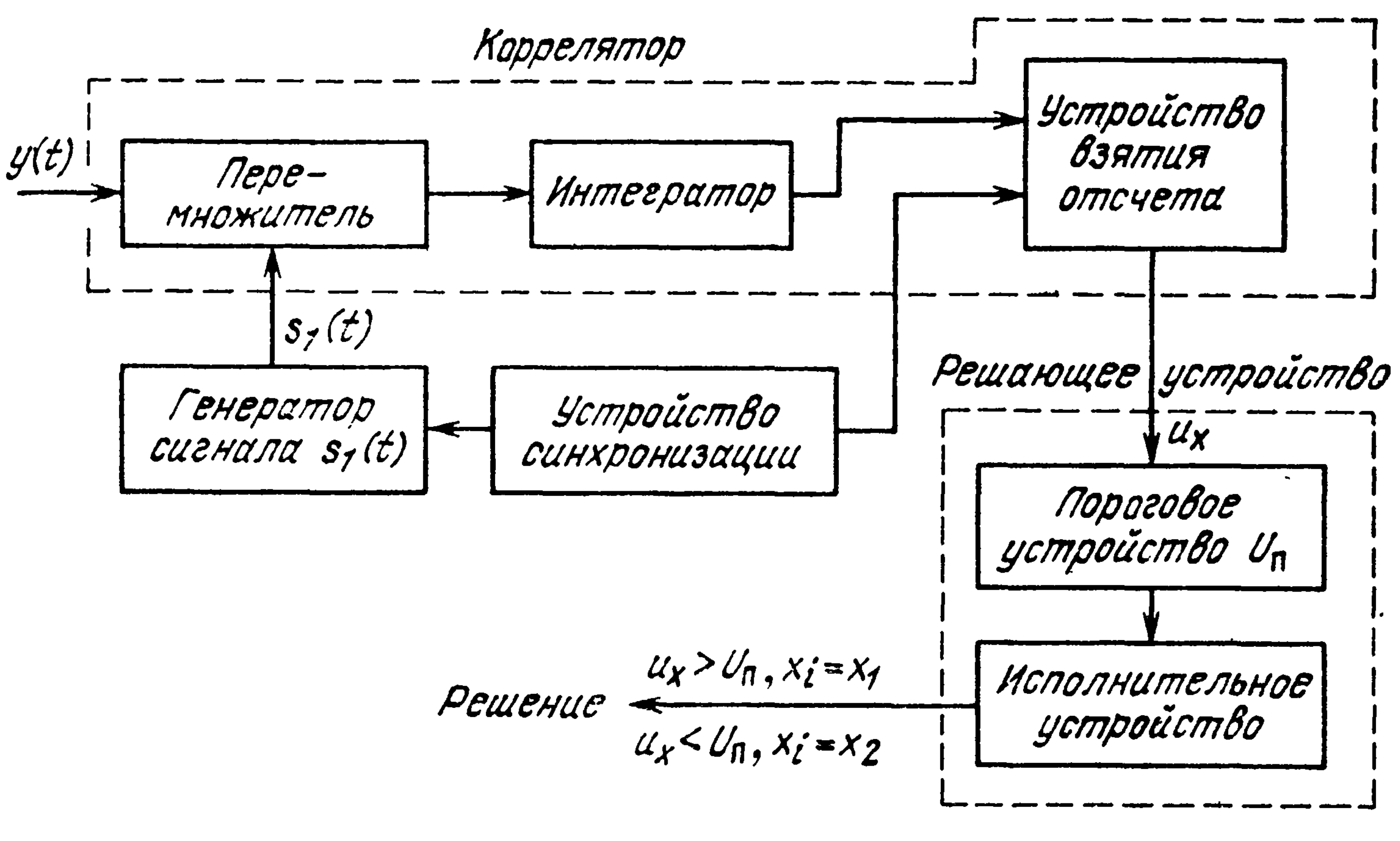


Рис. 3.3.

Каждый раз, когда анализ колебания $y(t)$ в корреляторе закончен, необходимо взять отсчет и полученную величину u_x подать в исполнительное устройство для принятия решения о том, какой символ был передан. Отсчеты берутся в моменты времени $t_k = k\tau_0$ ($k = 1, 2, \ldots$), соответствующие окончаниям информационных символов. Это достигается с помощью импульсов отсчета, вырабатываемых системой синхронизации. Таким образом, коррелятор состоит из перемножителя, интегратора и устройства взятия отсчета в моменты времени, кратные длительности информационного символа. Для обеспечения нормальной работы коррелятора в оптимальном приемнике необходимо располагать точной копией передаваемого сигнала $s_1(t)$.

Вся необходимая фильтрация сигнала, принятого на фоне помех, выполняется в корреляторе по видеочастоте. Никакой предварительной фильтрации сигнала до коррелятора в данном случае не требуется. Следова-

тельно, оптимальный корреляционный приемник можно рассматривать как демодулятор, состоящий из коррелятора и решающего устройства.

Заметим, что в любом приемнике, в том числе и оптимальном, результат обработки сигнала с шумом представляет собой некоторое случайное напряжение. Можно показать, что в рассматриваемом случае на выходе коррелятора образуется напряжение u_x, которое зависит от энергии принятого сигнала E и спектральной плотности белого шума N_0. Это напряжение сравнивается с напряжением порога $U_п$, и в соответствии с результатом сравнения автоматически принимается решение

$$u_x \gtrless U_п \quad \begin{matrix} \text{принят сигнал } s_1, \\ \text{принят сигнал } s_2. \end{matrix} \tag{3.37}$$

Рассмотренный способ приема часто называют когерентным оптимальным приемом, поскольку для перемножения используется опорное напряжение, полностью совпадающее (с точностью до постоянного множителя) с передаваемым сигналом.

Алгоритм (3.36) может быть реализован и иным способом. Как известно, напряжение на выходе инвариантного во времени линейного фильтра определяется выражением [3, 4]

$$u_ф(\tau) = \int_0^\tau y(t)\, W_ф(\tau - t)\, dt = \int_0^\tau W(t)\, y(\tau - t)\, dt, \tag{3.38}$$

где $y(t)$ — входное воздействие; $W(t)$ — импульсная переходная функция фильтра.

Если входной процесс $y(t)$ определен на интервале $0 \leqslant t \leqslant \tau_0$, напряжение на выходе фильтра в момент $\tau = \tau_0$ равно

$$u_ф(\tau)\big|_{\tau=\tau_0} = \int_0^{\tau_0} y(\tau_0 - t)\, W(t)\, dt. \tag{3.39}$$

Пусть фильтр согласован с посылкой сигнала $s_i(t)$ длительностью τ_0. При белом шуме импульсная переходная функция такого фильтра должна иметь вид [3, 4]

$$W_{фi}(t) = k s_i(\tau_0 - t), \tag{3.40}$$

где k — постоянный коэффициент, равный максимальному усилению фильтра. Не нарушая общности рассмотрения, в дальнейшем удобно положить $k=1$.

С учетом (3.40) выражение (3.39) принимает вид

$$u_{\text{ф}i}(\tau_0)=\int_0^{\tau_0} y(\tau_0-t)\,s_i(\tau_0-t)\,dt. \tag{3.41}$$

Нетрудно видеть, что после замены переменной $\tau_0-t=\zeta$ это выражение можно записать как

$$u_{\text{ф}i}(\tau_0)=\int_0^{\tau_0} y(\zeta)\,s_i(\zeta)\,d\zeta. \tag{3.42}$$

Полученный линейный функционал совпадает с корреляционным интегралом (3.31). Следовательно, алгоритм (3.36) можно реализовать с помощью линейного

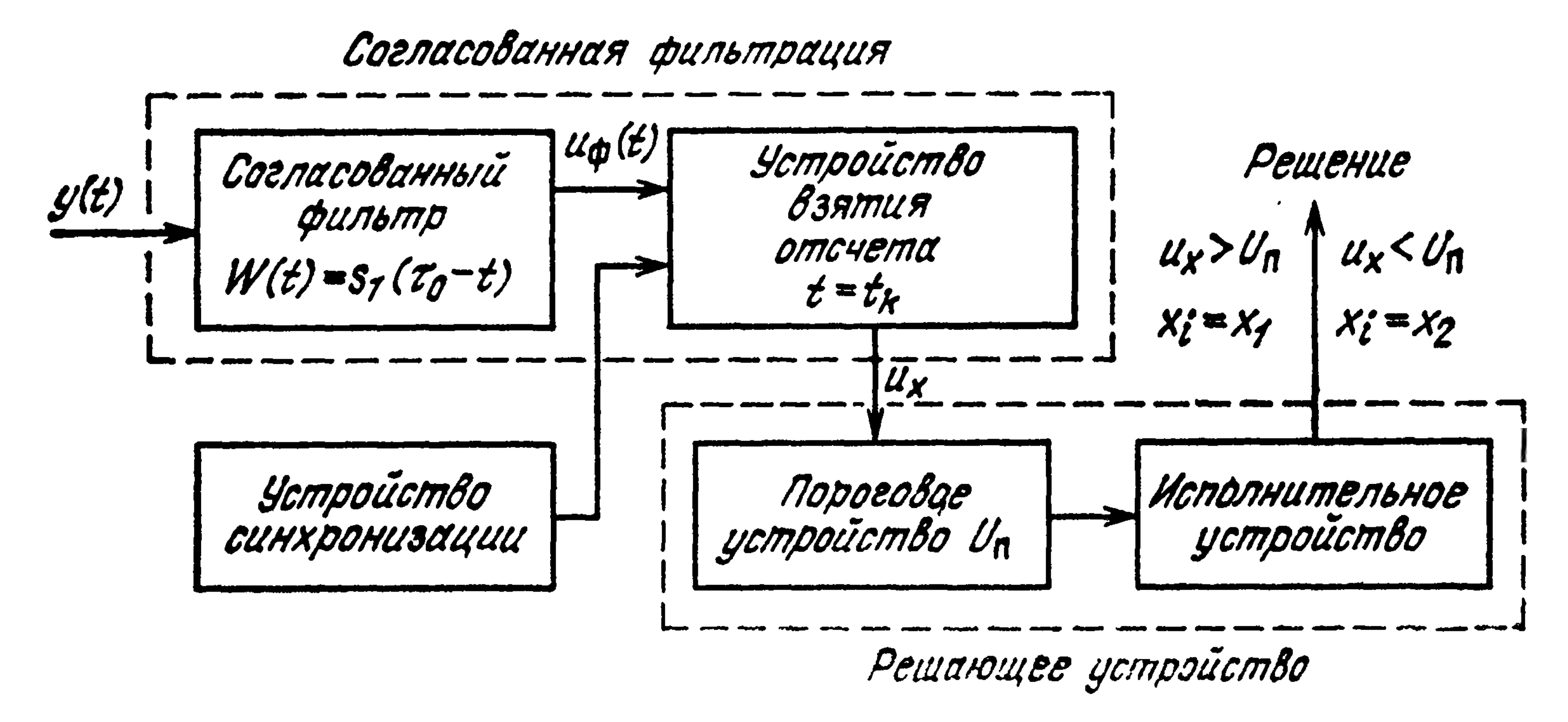


Рис. 3.4.

фильтра, согласованного с сигналом. Для этого необходимо брать отсчеты процесса на выходе фильтра в моменты времени t_k, соответствующие окончаниям информационных символов, сравнивать значения полученных отсчетов u_x с порогом $U_\text{п}$ и принимать решение о наличии того или иного сигнала на входе фильтра.

Структура соответствующего оптимального приемника приведена на рис. 3.4. В таком оптимальном приемнике вся необходимая фильтрация сигнала, принятого на фоне помех, выполняется согласованным фильтром на высокой частоте. Никакой специальной демодуляции сигнала, прошедшего через согласованный фильтр, не требуется.

Перейдем к рассмотрению второго случая.

3.3.2. Оптимальные приемники сигналов с активной паузой

Сигналом с активной паузой называют такой сигнал, который излучается при передаче любого информационного символа, будь то 1 или 0. Очевидно, что такой сигнал должен принимать два различных значения. Наиболее часто используют сигналы, для которых

$$E_1 = E_2 = E. \qquad (3.43)$$

Правило оптимального различения (3.33) в этом случае принимает вид

$$\zeta(s_1) - \zeta(s_2) \gtrless 0 \quad \begin{matrix} \text{принят сигнал } s_1, \\ \text{принят сигнал } s_2. \end{matrix} \qquad (3.44)$$

С учетом (3.31) это выражение можно записать в двух вариантах

$$1) \int_0^{\tau_0} y(t)\, s_1(t)\, dt - \int_0^{\tau_0} y(t)\, s_2(t)\, dt \gtrless 0 \quad \begin{matrix} \text{принят сигнал } s_1, \\ \text{принят сигнал } s_2; \end{matrix} \qquad (3.45)$$

$$2) \int_0^{\tau_0} y(t)\, \Delta s_{12}(t)\, dt \gtrless 0 \quad \begin{matrix} \text{принят сигнал } s_1, \\ \text{принят сигнал } s_2, \end{matrix} \qquad (3.46)$$

где

$$\Delta s_{12}(t) = s_1(t) - s_2(t). \qquad (3.47)$$

Оба варианта имеют определенный физический смысл и могут рассматриваться как разные алгоритмы работы оптимального приемника сигналов с активной паузой. В соответствии с алгоритмом (3.45) в оптимальном приемнике необходимо определять два корреляционных интеграла, брать их разность и отсчет этой разности в конце каждого интервала анализа τ_0 сравнивать с нулевым значением порога. Если отсчет больше нуля, принимается решение, что был передан символ x_1, если отсчет меньше нуля, — символ x_2. Подобная процедура может быть выполнена оптимальным приемником с двумя корреляторами либо с двумя согласованными фильтрами. Соответствующие схемы оптимальных приемников различения сигналов с активной паузой показаны на рис. 3.5 и 3.6.

При втором варианте алгоритма работы (3.46) оптимальный приемник должен вычислять один корреляционный интеграл и сравнивать его значение в конце

каждого интервала анализа τ_0 с нулевым значением порога. Так же, как и в предыдущих случаях, подобная процедура может быть выполнена либо приемником с коррелятором, либо приемником с согласованным фильтром. Схемы оптимальных приемников различения

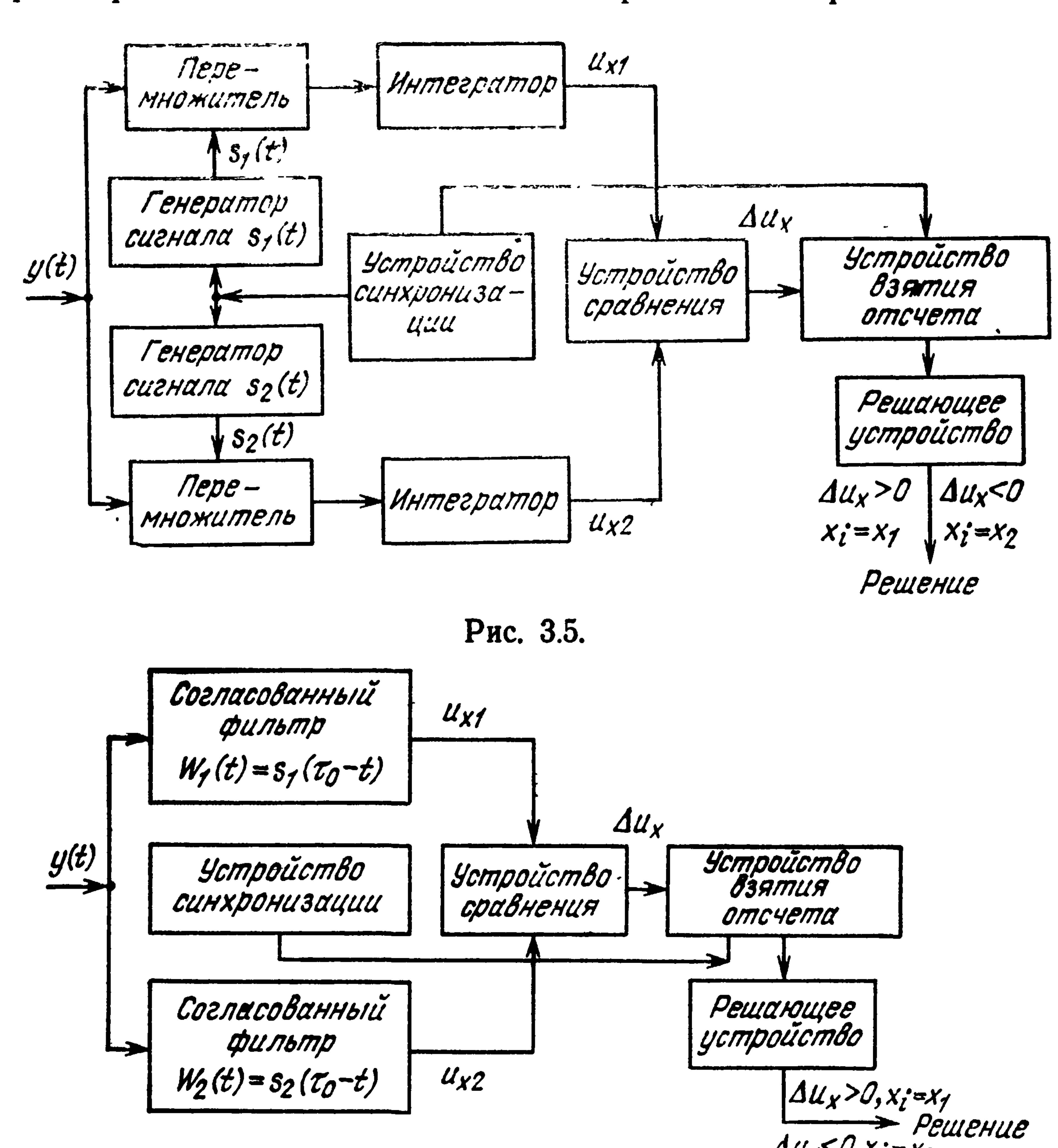


Рис. 3.5.

Рис. 3.6.

двух сигналов с активной паузой, соответствующие алгоритму (3.46), приведены на рис. 3.7 и 3.8. Их структура получается проще, чем структура оптимальных приемников, соответствующих алгоритму (3.45).

Проведенное рассмотрение показывает, что оптимальные приемники различения двух известных точно сигналов могут иметь различную структуру, которая

обеспечивает выполнение операций по анализу принятого колебания $y(t)$ и принятию решения о том, какой сигнал был передан. Однако в основе всех приемников лежит либо корреляционный способ обработки принятого колебания $y(t)$, либо оптимальная фильтрация с помощью линейных фильтров, согласованных с сигналами.

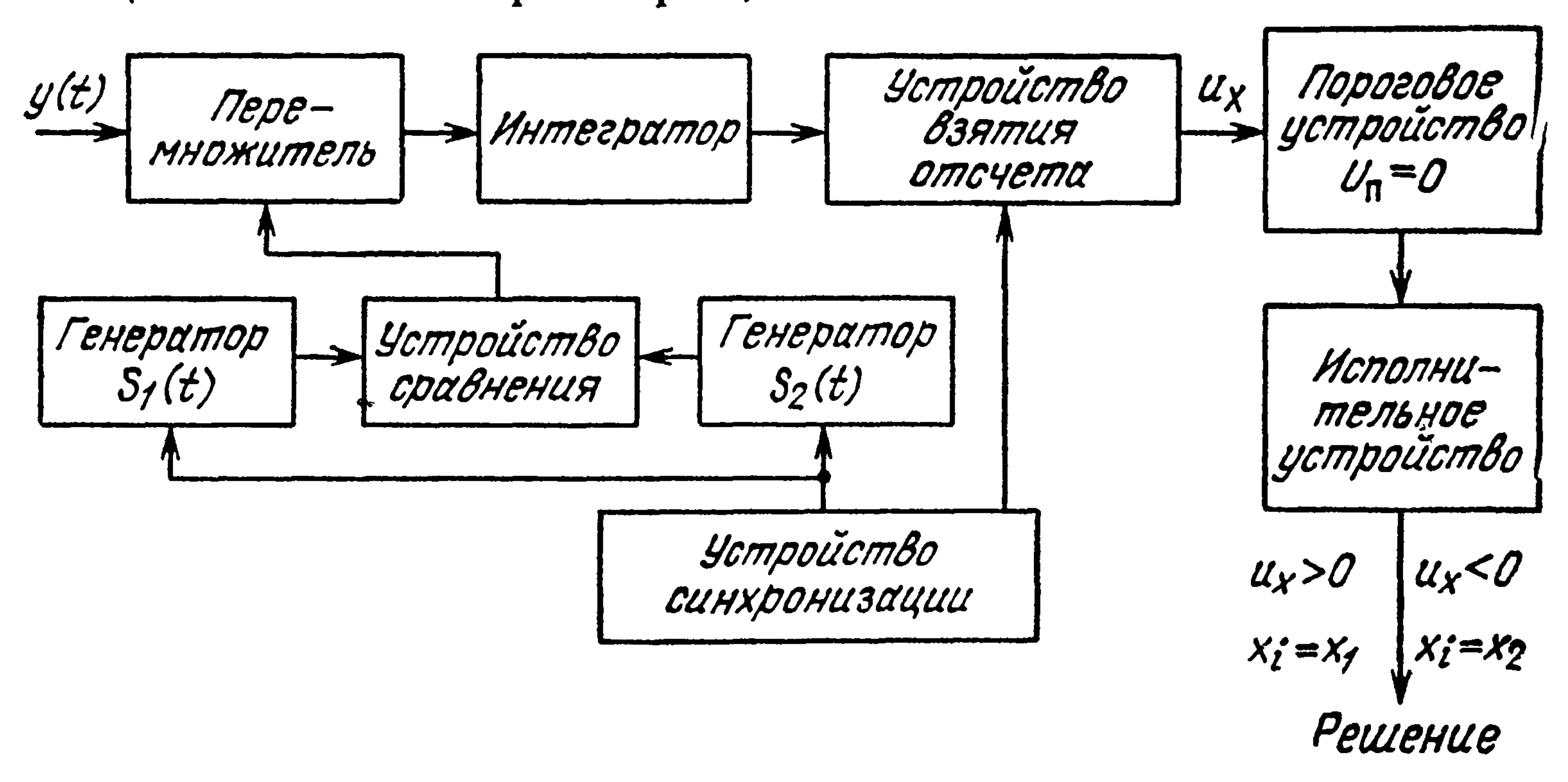


Рис. 3.7.

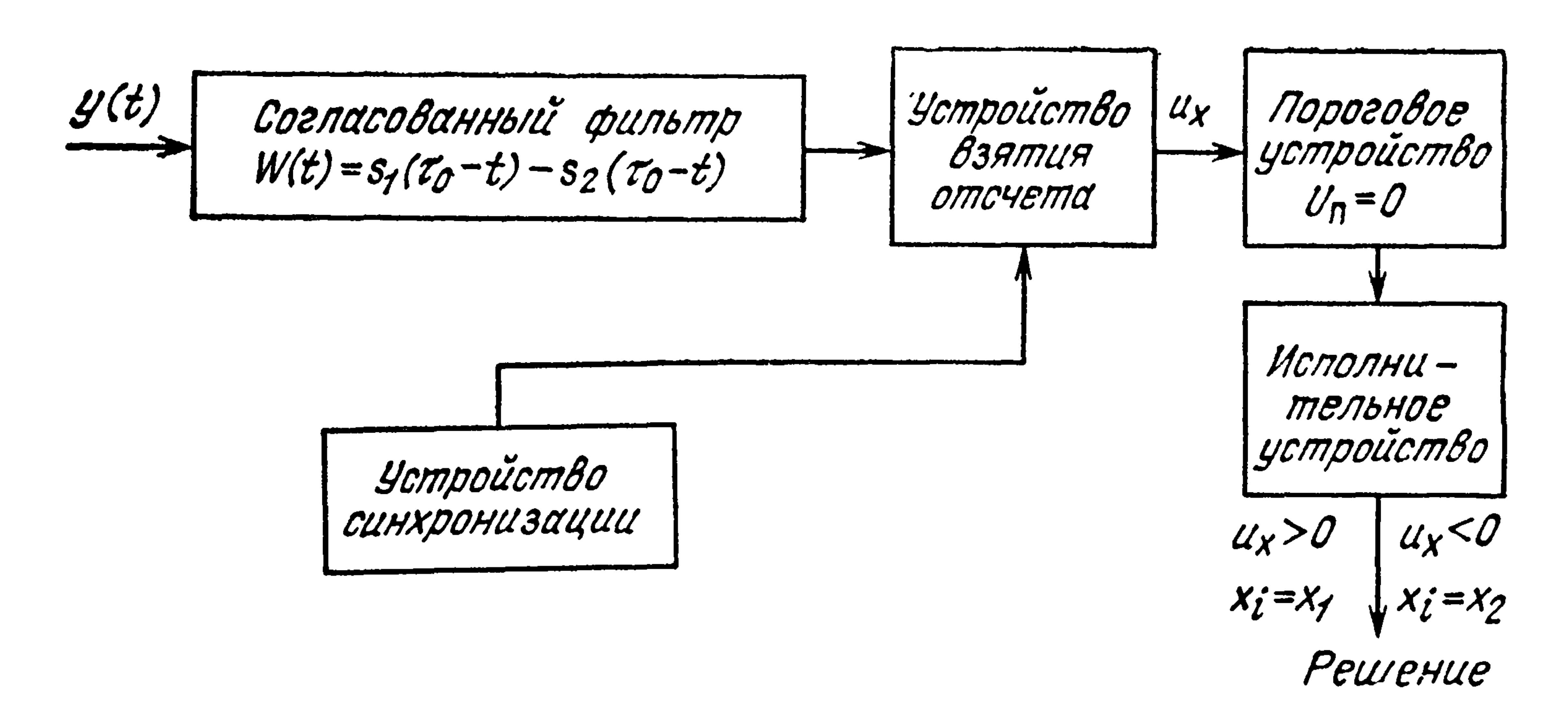


Рис. 3.8.

Итак, решение задачи различения на фоне белого шума двух известных точно сигналов позволило с помощью байесовского правила установить структуру оптимальных приемников, т. е. решить задачу синтеза оптимальной структуры приемной части цифровой системы связи*. Но это только часть задачи оптимизации цифровой системы связи.

* Здесь полезно вспомнить материал § 1.6.

Следующим важным вопросом является определение качества работы оптимальных приемников. Как уже указывалось выше, будем определять его величиной вероятности ошибки при передаче одного информационного символа.

3.4. ПОМЕХОУСТОЙЧИВОСТЬ ОПТИМАЛЬНЫХ ПРИЕМНИКОВ РАЗЛИЧЕНИЯ ДВУХ СИГНАЛОВ

При различении двух сигналов на фоне белого гауссовского шума оптимальный приемник кроме правильных решений иногда может принимать и ошибочные решения, обусловленные мешающим действием шума. Варианты возможных решений показаны в табл. 3.2.

Таблица 3.2

Переданный сигнал	Принятое решение	
$s(t) = s_1(t)$ $(x = x_1)$	$s(t) = s_1(t)$ $(\hat{x} = x_1)$ Правильно	$s(t) = s_2(t)$ $(\hat{x} = x_2)$ Ошибочно
$s(t) = s_2(t)$ $(x = x_2)$	$s(t) = s_2(t)$ $(\hat{x} = x_2)$ Правильно	$s(t) = s_1(t)$ $(\hat{x} = x_1)$ Ошибочно

Задача состоит в том, чтобы найти вероятности правильных и ошибочных решений. Введем следующие обозначения: $p(s_1|s_1)$ — вероятность принятия решения, что входное колебание $y(t)$ содержит сигнал $s_1(t)$ при условии, что был передан сигнал $s_1(t)$, т. е. вероятность правильного приема сигнала $s_1(t)$; $p(s_2|s_1)$ — вероятность принятия решения, что входное колебание $y(t)$ содержит сигнал $s_2(t)$, при условии, что был передан сигнал $s_1(t)$, т. е. вероятность ошибочного приема сигнала $s_1(t)$.

Аналогично определяются и вероятности приема сигнала $s_2(t)$: $p(s_2|s_2)$ — вероятность правильного прие-

ма сигнала $s_2(t)$; $p(s_1|s_2)$ — вероятность ошибочного приема сигнала $s_2(t)$. Нетрудно видеть, что рассмотренные вероятности удовлетворяют условиям

$$\begin{aligned} p(s_1|s_1)+p(s_2|s_1)&=1, \\ p(s_2|s_2)+p(s_1|s_2)&=1. \end{aligned} \qquad (3.48)$$

Чтобы определить качество работы оптимального приемника, достаточно знать вероятности ошибочного приема сигналов $s_1(t)$ и $s_2(t)$, т. е. вероятности $p(s_2|s_1)$ и $p(s_1|s_2)$. Если априорные вероятности $p(s_1)$ и $p(s_2)$ появления сигналов $s_1(t)$ и $s_2(t)$ на интервале анализа $(0-\tau_0)$ известны, то вероятность ошибки при анализе колебания $y(t)$ на интервале $(0-\tau_0)$ равна

$$p_{\text{ош}}=p(s_1)p(s_2|s_1)+p(s_2)p(s_1|s_2). \qquad (3.49)$$

Поскольку между сигналами $s_1(t)$ и $s_2(t)$ и информационными символами x_1 и x_2 имеется однозначное соответствие, то выражение (3.49) можно записать и в другом виде:

$$p_{\text{ош}}=p(x_1)p(x_2|x_1)+p(x_2)p(x_1|x_2), \qquad (3.50)$$

где $p(x_1)$ и $p(x_2)$ — априорные вероятности передачи символов x_1 и x_2; $p(x_2|x_1)$ — вероятность приема символа x_2 при условии, что был передан символ x_1; $p(x_1|x_2)$ — вероятность приема символа x_1 при условии, что был передан символ x_2. Величина $p_{\text{ош}}$ характеризует вероятность ошибки при передаче одного информационного символа по каналу связи с помехами.

Как уже неоднократно указывалось, для двоичных цифровых последовательностей можно считать справедливым условие $p(s_1)=p(s_2)=0{,}5$, при котором выражения (3.49) и (3.50) принимают вид

$$p_{\text{ош}}=0{,}5[p(s_2|s_1)+p(s_1|s_2)], \qquad (3.51)$$

$$p_{\text{ош}}=0{,}5[p(x_2|x_1)+p(x_1|x_2)]. \qquad (3.52)$$

Если $p(s_2|s_1)=p(s_1|s_2)$, то

$$p_{\text{ош}}=p(s_1|s_2)=p(s_2|s_1). \qquad (3.53)$$

Канал, в котором условные вероятности ошибок равны между собой, называется симметричным.

3.4.1. Вероятность ошибки при оптимальном приеме сигналов с пассивной паузой

Предположим, что передан сигнал $s_1(t)$. Тогда колебание $y(t)$, поступившее на вход оптимального приемника, имеет вид

$$y(t)=s_1(t)+n(t).$$

В этом случае приемник должен вычислить значение корреляционного интеграла

$$\zeta_1=\zeta(s_1)=\int_0^{\tau_0}[s_1(t)+n(t)]\,s_1(t)\,dt, \qquad (3.54)$$

сравнить его с величиной порога $q=E/2$ и принять решение в соответствии с правилом (3.35), которое в рассматриваемом случае приобретает вид

$$\zeta_1 \gtrless q \quad \begin{matrix}\text{принят сигнал } s_1,\\ \text{принят сигнал } s_2.\end{matrix} \qquad (3.55)$$

Величина ζ_1 является случайной, так как зависит не только от сигнала, но и от шума. Поэтому условие $\zeta_1>q$ (или $\zeta_1<q$) является статистическим и выполняется с некоторой вероятностью. Вероятность выполнения условия $\zeta_1>q$ есть вероятность правильного приема сигнала $s_1(t)$, а вероятность невыполнения этого условия — вероятность ошибочного приема сигнала $s_1(t)$.

Итак если передан сигнал $s_1(t)$, то вероятность его правильного приема

$$p(s_1|s_1)=p(\zeta_1\geqslant q), \qquad (3.56)$$

а вероятность ошибочного приема

$$p(s_2|s_1)=p(\zeta_1<q)=1-p(\zeta_1\geqslant q). \qquad (3.57)$$

Если передан сигнал $s_2(t)$, колебание $y(t)$ на входе приемника определяется только шумом, т. е. $y(t)=n(t)$. Значение корреляционного интервала при этом определяется величиной

$$\zeta_2=\zeta(s_2)=\int_0^{\tau_2}n(t)\,s_1(t)\,dt, \qquad (3.58)$$

а правило решения принимает вид

$$\zeta_2 \lessgtr q \quad \begin{matrix}\text{принят сигнал } s_2,\\ \text{принят сигнал } s_1.\end{matrix} \qquad (3.59)$$

В соответствии с этим правилом условная вероятность правильного приема сигнала $s_2(t)$ определяется вероятностью выполнения условия $\zeta_2 < q$:

$$p(s_2|s_2) = p(\zeta_2 < q), \tag{3.60}$$

а вероятность ошибочного приема — вероятностью невыполнения этого условия:

$$p(s_1|s_2) = p(\zeta_2 \geqslant q) = 1 - p(\zeta_2 < q). \tag{3.61}$$

Чтобы конкретизировать все эти вероятности, необходимо знать законы распределения случайных величин ζ_1 и ζ_2 и параметры этих распределений.

Из выражений (3.54) и (3.58) следует, что случайные величины ζ_1 и ζ_2 являются результатом линейного преобразования нормального белого шума и, следовательно, имеют нормальный закон распределения*. Найдем параметры этого распределения для ζ_1 и ζ_2. Нетрудно видеть, что среднее значение (математическое ожидание) величины ζ_2 равно нулю. Действительно, имеем

$$\overline{\zeta_2} = \int_0^{\tau_0} \overline{n(t)\, s_1(t)}\, dt = \int_0^{\tau_0} \overline{n(t)}\; \overline{s_1(t)}\, dt = 0. \tag{3.62}$$

В этом выражении и далее черта означает среднее значение (математическое ожидание). Здесь учтено, что среднее значение произведения двух независимых величин равно произведению их средних, а среднее значение белого шума равно нулю $(\overline{n(t)} = 0)$.

Дисперсия случайной величины ζ определяется выражением

$$D(\zeta) = \overline{(\zeta - \overline{\zeta})^2}. \tag{3.63}$$

Учитывая выражение (3.58) и (3.62), для дисперсии ζ_2 имеем

$$D(\zeta_2) = \overline{\left(\int_0^{\tau_0} n(t)\, s_1(t)\, dt\right)^2} =$$

$$= \int_0^{\tau_0}\int_0^{\tau_0} \overline{n(t)\, n(u)\, s_1(t)\, s_1(u)}\, dt du.$$

---

* Подобные линейные преобразования в литературе часто называют линейными функционалами нормального белого шума [3, 8].

Так как сигнал и шум независимы, то

$$D(\zeta_2)=\int_0^{\tau_0}\int_0^{\tau_0} K_{ш}(t,\ u)\overline{s_1(t)\, s_1(u)}\, dtdu, \tag{3.64}$$

где

$$K_{ш}(t,\ u)=\overline{n(t)\, n(u)}=0{,}5N_0\delta(t-u)$$

— корреляционная функция белого шума с односторонней спектральной плотностью N_0 [2, 3].

Принимая во внимание фильтрующие свойства дельта-функции, окончательно имеем *

$$D(\zeta_2)=0{,}5N_0\int_0^{\tau_0}\overline{s^2{}_1(u)}\, du=N_0E/2. \tag{3.65}$$

Нетрудно показать, проведя аналогичные вычисления для величины ζ_1, определяемой выражением (3.54), что

$$\overline{\zeta}_1=E;\ D(\zeta_1)=N_0E/2. \tag{3.66}$$

Теперь, когда параметры законов распределения случайных величин ζ_1 и ζ_2 известны, можно записать одномерные нормальные распределения этих величин в виде **

$$W(\zeta_1)=\frac{1}{\sqrt{2\pi N_0E/2}}\exp\left[-\frac{(\zeta_1-E)^2}{N_0E}\right], \tag{3.67}$$

$$W(\zeta_2)=\frac{1}{\sqrt{2\pi N_0E/2}}\exp\left[-\frac{\zeta^2{}_2}{N_0E}\right]. \tag{3.68}$$

Эти распределения показаны на рис. 3.9.

Условные вероятности ошибочного приема сигналов $s_1(t)$ и $s_2(t)$ в соответствии с (3.57) и (3.61) определя-

* Под фильтрующими свойствами дельта-функции понимается свойство, которое может быть иллюстрировано выражением

$$\int_{t_1}^{t_2} f(t)\delta(t-\tau)dt=f(\tau),$$

показывающим, что значение интеграла равно значению функции $f(t)$, взятому в момент времени $t=\tau$. Подробнее со свойствами дельта-функции можно ознакомиться в работах [2, 7], см. также § 2.2.

** Законы распределения величин ζ_1 и ζ_2 являются одномерными, так как значения этих величин берутся в момент времени, соответствующий окончанию анализа входного колебания в оптимальном приемнике.

ются выражениями

$$p(s_2|s_1) = p(\zeta_1 < q) = \int\limits_{-\infty}^{q} W(\zeta_1)\, d\zeta_1, \tag{3.69}$$

$$p(s_1|s_2) = p(\zeta_2 > q) = \int\limits_{q}^{\infty} W(\zeta_2)\, d\zeta_2. \tag{3.70}$$

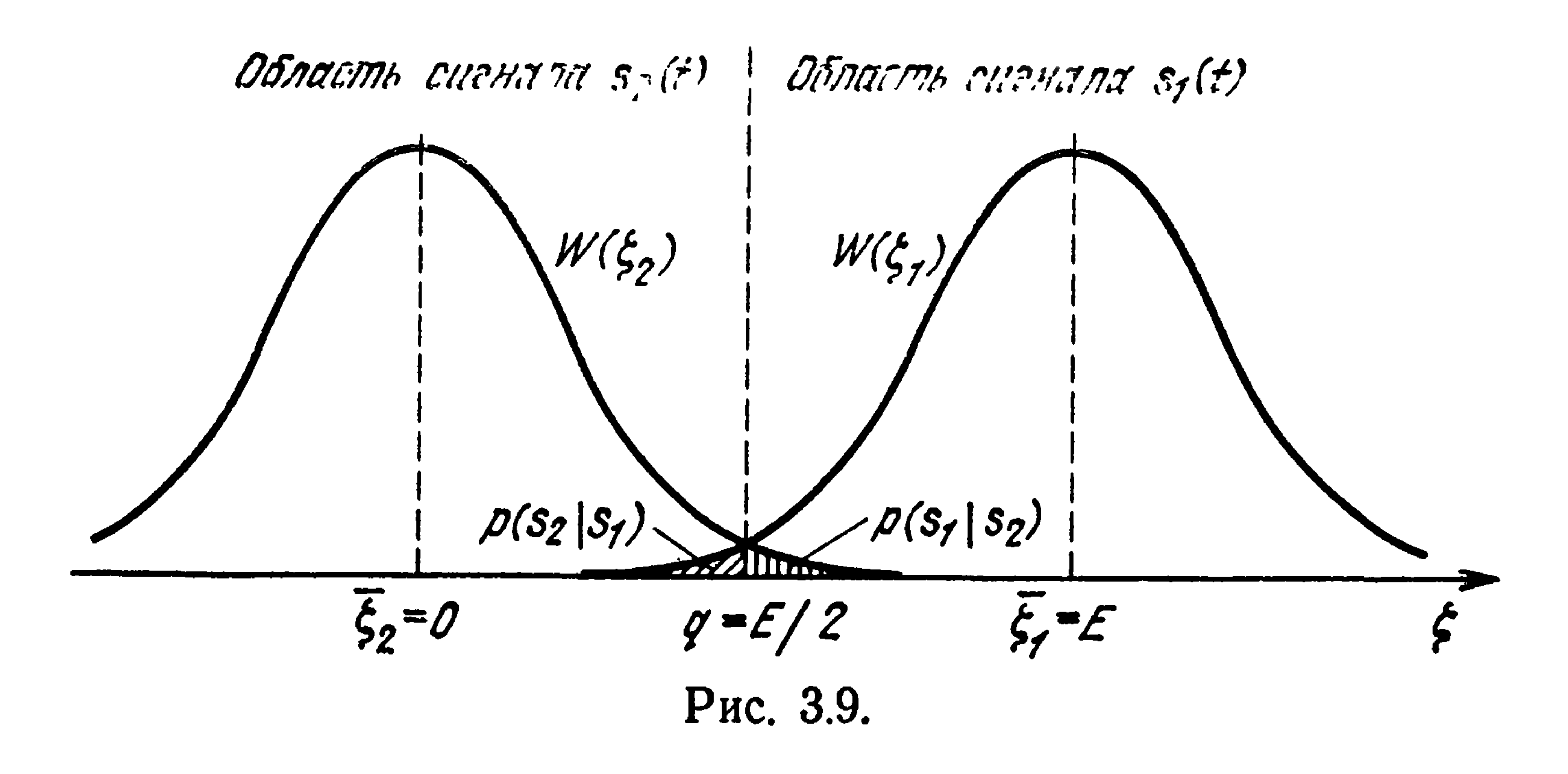


Рис. 3.9.

Несложно показать, что с учетом (3.67), (3.68) и после замены переменных эти выражения можно привести к виду

$$p(s_2|s_1) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{-\sqrt{E/2N_0}} \exp\left(-\frac{y^2}{2}\right) dy, \tag{3.71}$$

где новая переменная

$$y = (\zeta_1 - E)/\sqrt{N_0 E/2},$$

$$p(s_1|s_2) = \frac{1}{\sqrt{2\pi}} \int\limits_{\sqrt{E/2N_0}}^{\infty} \exp\left(-\frac{y^2}{2}\right) dy, \tag{3.72}$$

где новая переменная

$$y = \zeta_2/\sqrt{N\ E/2}.$$

Из рассмотрения этих выражений следует, что

$$p(s_2|s_1) = p(s_1|s_2) = V(\sqrt{E/2N_0}), \tag{3.73}$$

где

$$V(\alpha)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{-\alpha}\exp\left(-\frac{y^2}{2}\right)dy \qquad (3.74)$$

— интеграл ошибок.

Равенство условных вероятностей ошибок в рассматриваемом случае объясняется тем, что функций распределения $W(\zeta_1)$ и $W(\zeta_2)$ симметричны относительно порогового значения q (см. рис. 3.9).

Учитывая выражения (3.53) и (3.73), получаем формулу для определения вероятности ошибки при оптимальном приеме сигналов с пассивной паузой

$$p_{\text{ош}}=p(s_2|s_1)=p(s_1|s_2)=V(\sqrt{E/2N_0}). \qquad (3.75)$$

Для записи вероятностей ошибок в литературе широко используется большое число разновидностей функций, связанных с интегралом ошибок. Все они различаются между собой коэффициентами и пределами интегрирования.

Наиболее часто применяются функции

$$F(\alpha)=1-V(\alpha)=\frac{1}{\sqrt{2\pi}}\int\limits_{-\infty}^{\alpha}\exp\left(\frac{-y^2}{2}\right)dy, \qquad (3.76)$$

называемая нормальной функцией распределения, и

$$\Phi(\alpha)=0{,}5-V(\alpha)=\frac{1}{\sqrt{2\pi}}\int\limits_{0}^{\alpha}\exp\left(\frac{-y^2}{2}\right)dy, \qquad (3.77)$$

называемая интегралом вероятностей или функцией Лапласса.

В математической литературе широкое применение находят функции

$$\operatorname{erf}(\alpha)=\frac{2}{\sqrt{\pi}}\int\limits_{0}^{\alpha}\exp(-y^2)\,dy \text{ и } \operatorname{erfc}(\alpha)=1-\operatorname{erf}(\alpha).$$

Нетрудно видеть, что

$$\operatorname{erf}(\alpha)=2F(\sqrt{2}\,\alpha)-1 \text{ или } \operatorname{erf}(\alpha)=2\Phi(\sqrt{2}\,\alpha).$$

Выбор той или иной функции часто зависит только от вкуса и привычек автора. Поэтому при чтении специальной литературы во избежание ошибок и недоразумений

необходимо тщательно уяснить, какими функциями пользуется данный автор.

Все указанные выше функции табулированы в ряде книг и справочников. В частности, таблицы для $F(\alpha)$ имеются в [3, 9], а для $\Phi(\alpha)$ — в [10, 11]. Краткая таблица для $V(\alpha)$ дана в приложении 1.

В дальнейшем, как правило, мы будем пользоваться функцией Лапласа (3.77). В этом случае выражение (3.75) для вероятности ошибки принимает вид

$$p_{\text{ош}} = 0{,}5 - \Phi(\sqrt{E/2N_0}). \qquad (3.78)$$

Обсудим полученный результат. Формула (3.78) соответствует условию равных априорных вероятностей передачи сигналов $s_1(t)$ и $s_2(t)$, а следовательно, и информационных символов x_1 и x_2. С точки зрения теории информации это означает, что каждый переданный символ содержит в среднем одну двоичную единицу информации (один бит). Значит, формулу (3.78) можно рассматривать как вероятность ошибки при передаче одной двоичной единицы информации по гауссовскому каналу связи сигналами с пассивной паузой и оптимальном приеме. Эта вероятность полностью определяется величиной энергетического отношения сигнала к шуму. Введем обозначение

$$h^2_0 = E/N_0. \qquad (3.79)$$

Тогда

$$p_{\text{ош}} = 0{,}5 - \Phi(h_0/\sqrt{2}). \qquad (3.80)$$

Дальнейшее обсуждение проведем после рассмотрения следующего случая.

3.4.2. Вероятность ошибки при оптимальном приеме сигналов с активной паузой

Ограничим рассмотрение наиболее важным и часто встречающимся условием равенства энергий сигналов $s_1(t)$ и $s_2(t)$ и равенством априорных вероятностей передачи этих сигналов, т. е.

$$E_1 = E_2 = E \text{ и } p(s_1) = p(s_2). \qquad (3.81)$$

Предположим, что был послан сигнал $s_1(t)$. Следовательно, на вход приемника поступит колебание $y(t) =$

$=s_1(t)+n(t)$. Тогда в соответствии с правилом (3.46) оптимальный приемник должен определить величину

$$\zeta_1=\int_0^{\tau_0}[s_1(t)+n(t)][s_1(t)-s_2(t)]\,dt \tag{3.82}$$

и сравнить ее с нулевым значением порога. Если $\zeta_1>0$, то сигнал $s_1(t)$ будет принят правильно, если же $\zeta_1<0$, то произойдет ошибка и вместо сигнала $s_1(t)$ будет принят сигнал $s_2(t)$.

Исследуем случайную величину ζ_1, чтобы найти ее закон распределения и параметры этого распределения. Развернем выражение (3.82):

$$\zeta_1=\int_0^{\tau_0}s^2{}_1(t)\,dt-\int_0^{\tau_0}s_1(t)\,s_2(t)\,dt+\int_0^{\tau_0}n(t)[s_1(t)-s_2(t)]\,dt. \tag{3.83}$$

В этом выражении первое слагаемое представляет собой энергию сигнала $s_1(t)$, а второе — функцию взаимной корреляции между сигналами $s_1(t)$ и $s_2(t)$. Обозначим ее

$$\varphi_s=\int_0^{\tau_0}s_1(t)\,s_2(t)\,dt \tag{3.84}$$

и введем коэффициент взаимной корреляции между сигналами

$$\rho_s=\varphi_s/E=\int_0^{\tau_0}s_1(t)\,s_2(t)\,dt\Bigg/\int_0^{\tau_0}s^2{}_1(t)\,dt. \tag{3.85}$$

Тогда получим

$$\zeta_1=E(1-\rho_s)+\int_0^{\tau_0}n(t)[s_1(t)-s_2(t)]\,dt. \tag{3.86}$$

Из этого выражения следует, что случайная величина ζ_1 (так же, как и в предыдущем случае) подчиняется нормальному закону распределения. Опираясь на результаты рассмотрения предыдущего случая, нетрудно показать, что среднее значение величины ζ_1

$$\overline{\zeta}_1=E(1-\rho_s), \tag{3.87}$$

а дисперсия

$$D(\zeta_1)=N_0E(1-\rho_s). \tag{3.88}$$

Аналогичное рассмотрение можно провести и для величины ζ_2 Пусть посылался сигнал $s_2(t)$ Тогда на вход приемника поступит колебание $y(t)=s_2(t)+n(t)$. В соответствии с правилом (3 46) должна быть определена величина

$$\zeta_2 = \int_0^{\tau_0} [s_2(t)+n(t)][s_1(t)-s_2(t)]\,dt \tag{3.89}$$

и проведено ее сравнение с нулевым значением порога Если $\zeta_2 \geqslant 0$, произойдет ошибка и вместо сигнала $s_2(t)$ будет принят сигнал $s_1(t)$, если же $\zeta_2<0$, то сигнал $s_2(t)$ будет принят правильно

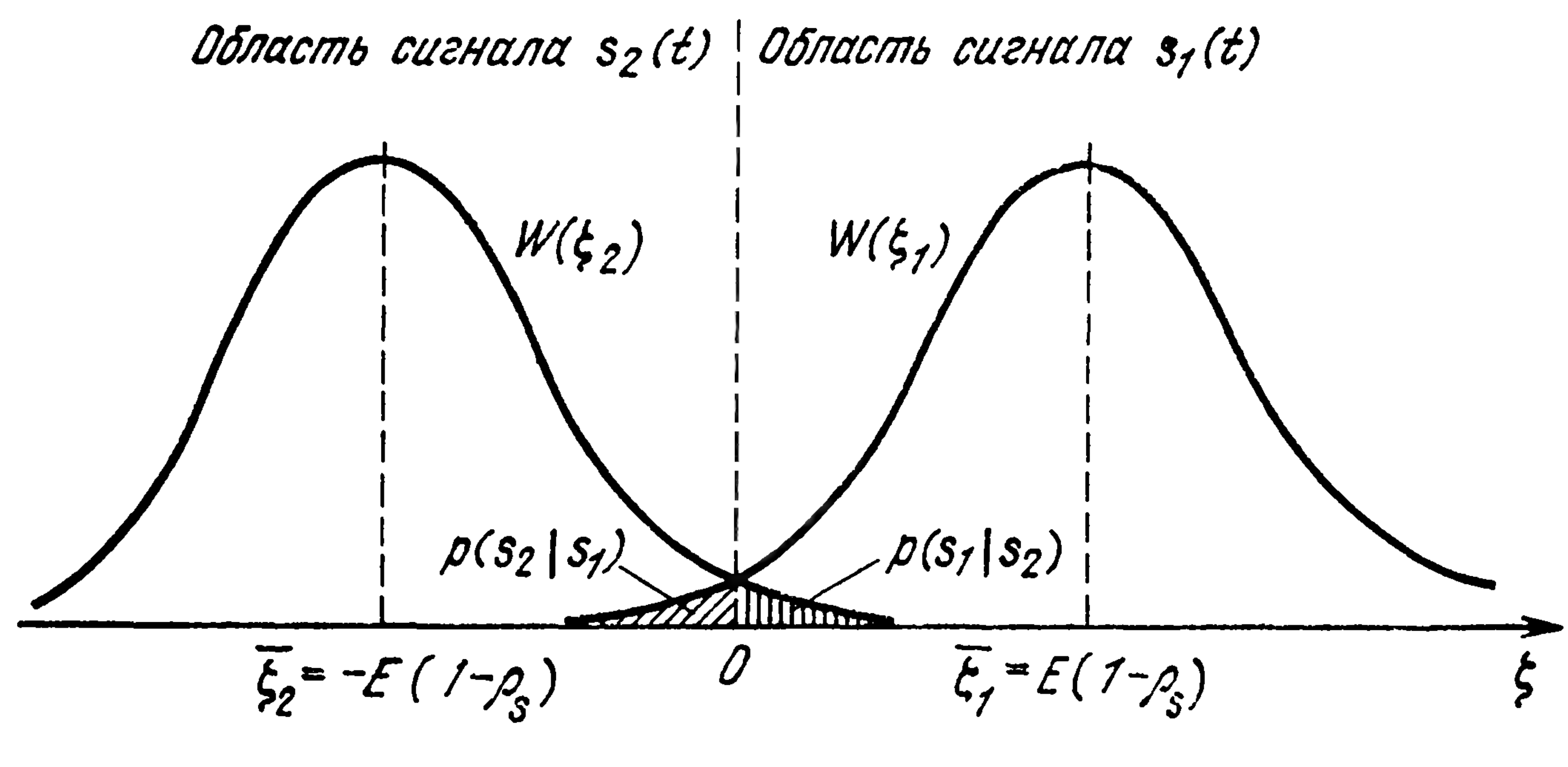


Рис 3 10

Проанализировав выражение для ζ_2, нетрудно убедиться, что величина ζ_2 так же, как и ζ_1, распределена по нормальному закону, а ее среднее значение и дисперсия определяются выражениями

$$\bar{\zeta}_2 = -E(1-\rho_s), \tag{3.90}$$

$$D(\zeta_2) = N_0 E(1-\rho_s). \tag{3.91}$$

С учетом полученных параметров законы распределения случайных величин ζ_1 и ζ_2 можно записать в виде

$$W(\zeta_1) = \frac{1}{\sqrt{2\pi N_0 E(1-\rho_s)}} \exp\left\{-\frac{[\zeta_1 - E(1-\rho_s)]^2}{2N_0E(1-\rho_s)}\right\}, \tag{3.92}$$

$$W(\zeta_2) = \frac{1}{\sqrt{2\pi N_0 E(1-\rho_s)}} \exp\left\{\frac{[\zeta_2 + E(1-\rho_s)]^2}{2N_0E(1-\rho_s)}\right\}. \tag{3.93}$$

Эти распределения показаны на рис. 3 10.

Условные вероятности ошибочного приема сигналов s_1 и s_2 определяются выражениями

$$p(s_2|s_1) = p(\zeta_1 < 0) = \int_{-\infty}^{0} W(\zeta_1)\, d\zeta_1, \quad (3.94)$$

$$p(s_1|s_2) = p(\zeta_2 \geqslant 0) = \int_{0}^{\infty} W(\zeta_2)\, d\zeta_2. \quad (3.95)$$

Найдем вероятность $p(s_2|s_1)$. Подставив в (3 94) значение $W(\zeta_1)$ в соответствии с (3 92) и заменив переменную интегрирования

$$[\zeta_1 - E(1-\rho_s)]/\sqrt{N_0 E(1-\rho_s)} = y,$$

получим

$$p(s_2|s_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-\overline{\zeta}_1/\sqrt{D(\zeta_1)}} \exp\left(\frac{-y^2}{2}\right) dy =$$

$$= \Phi(\infty) - \Phi\left(\frac{\overline{\zeta_1}}{\sqrt{D(\zeta_1)}}\right),$$

где $\Phi(\alpha)$ определяется выражением (3.77).

Приняв во внимание выражения (3 87) и (3.88) для ζ_1 и $D(\zeta_1)$ и учтя, что $\Phi(\infty)=0{,}5$, имеем

$$p(s_2|s_1) = 0{,}5 - \Phi\left[\sqrt{\frac{E}{N_0}(1-\rho_s)}\right]. \quad (3.96)$$

Выполнив аналогичные преобразования, нетрудно убедиться в том, что вероятность $p(s_1|s_2)$ определяется точно таким же выражением Равенство вероятностей $p(s_1|s_2)$ и $p(s_2|s_1)$ объясняется тем, что распределения (3.92) и (3.93) имеют одинаковую дисперсию и симметричны относительно нулевого порогового уровня. Следовательно, в соответствии с выражениями (3.53) и (3.96) вероятность ошибки при передаче одного символа информации равна

$$p_{\text{ош}} = 0{,}5 - \Phi\left[\sqrt{\frac{E}{N_0}(1-\rho_s)}\right] \quad (3.97)$$

или с учетом (3.79)

$$p_{\text{ош}} = 0{,}5 - \Phi(h_0\sqrt{1-\rho_s}). \quad (3.98)$$

Так же, как и для сигналов с пассивной паузой, это выражение определяет вероятность ошибки при передаче одной двоичной единицы информации, поскольку при рассмотрении было принято $p(s_1) = p(s_2)$

3.5. СРАВНЕНИЕ ПОМЕХОУСТОЙЧИВОСТИ ОПТИМАЛЬНОГО ПРИЕМА ПРИ РАЗНЫХ ВИДАХ СИГНАЛОВ

Представим выражения (3 80) и (3.98) для сигналов с пассивной и активной паузами в универсальном виде·

$$p_{\text{ош}} = 0{,}5 - \Phi(\alpha), \qquad (3\,99)$$

где для сигналов с пассивной паузой

$$\alpha^2 = \alpha^2_{\text{п}} = E/2N_0 = h^2_0/2, \qquad (3.100)$$

а для сигналов с активной паузой

$$\alpha^2 = \alpha^2_{\text{а}} = \frac{E}{N_0}(1 - \rho_s) = h^2_0(1 - \rho_s). \qquad (3.101)$$

Зависимость $p_{\text{ош}} = f(\alpha)$ приведена на рис 3 11 Из рассмотрения графика этой зависимости следует, что для получения малой вероятности ошибки ($p_{\text{ош}} \leqslant 10^{-3}$) необходимо обеспечить величину $\alpha \geqslant 3$. Для таких значений α интеграл вероятности $\Phi(\alpha)$ можно разложить в асимптотический ряд и пользоваться его приближенным представлением в виде

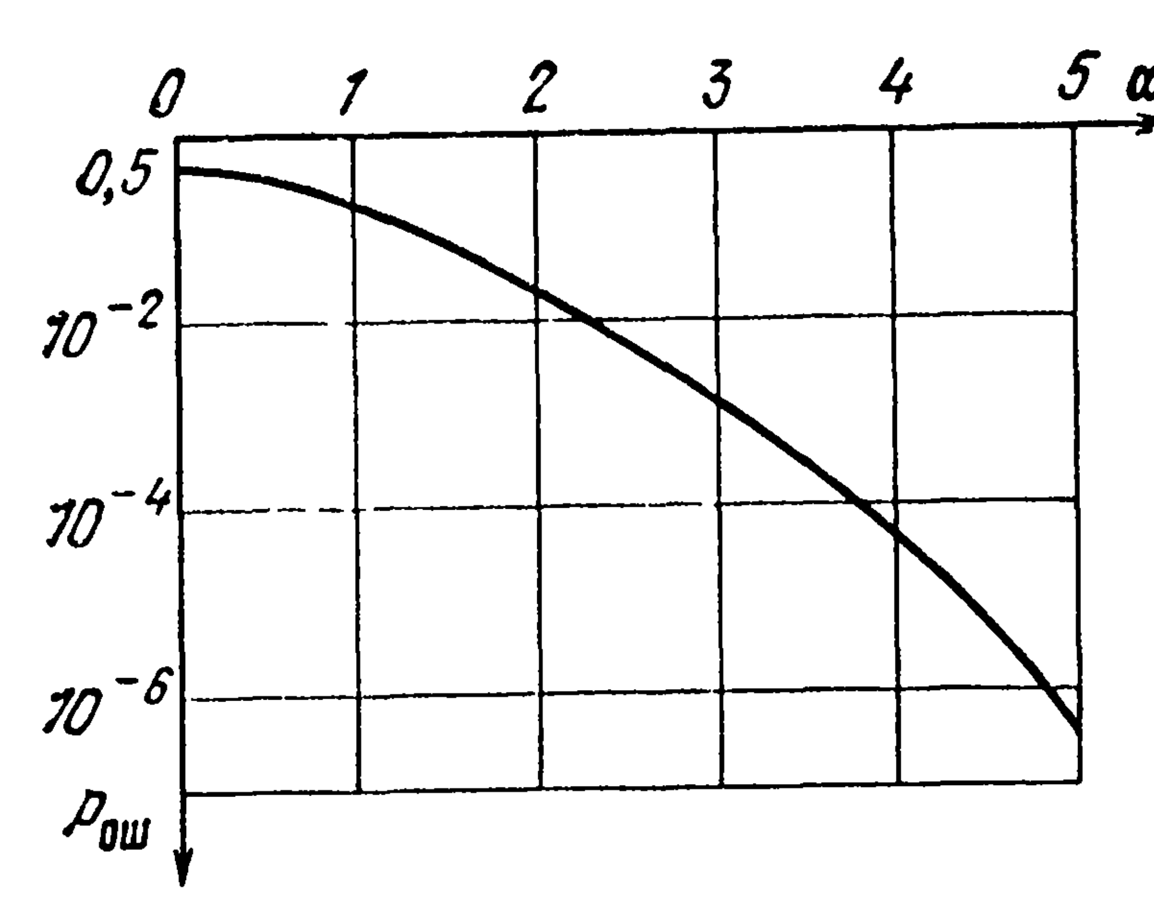


Рис. 3.11.

$$\Phi(\alpha) = 0{,}5 - \frac{\exp(-\alpha^2/2)}{\sqrt{2\pi}\alpha}\left[1 - \frac{1}{\alpha^2} + \ldots\right]. \qquad (3.102)$$

Тогда с точностью не хуже 10% выражение (3.99) можно представить в виде

$$p_{\text{ош}} \approx \frac{1}{\sqrt{2\pi}\,\alpha}\exp\left(-\frac{\alpha^2}{2}\right), \quad \alpha \geqslant 3. \qquad (3.103)$$

Во многих случаях, когда требования к помехоустойчивости достаточно высоки, удобнее пользоваться выражением (3.103), чем (3.99).

При заданной спектральной плотности шума N_0 необходимая величина α, при которой достигается нужная помехоустойчивость, для сигналов с пассивной паузой может быть получена только за счет энергии сигнала s_1 (поскольку $s_2 \equiv 0$).

Для сигналов с активной паузой необходимая величина α зависит не только от энергии сигналов s_1 и s_2, но и от величины коэффициента взаимной корреляции ρ_s между этими сигналами, т. е. от конкретного вида этих сигналов. В зависимости от вида сигналов величина ρ_s (см. выражение (3.85)) может принимать различные значения в интервале $-1 \leqslant \rho_s \leqslant +1$. Величина взаимной корреляции характеризует различие между сигналами, поэтому значение ρ_s можно рассматривать как меру этого различия, а границы возможных значений ρ_s — как пределы этой меры.

Выделим три наиболее характерных случая.

1. $\rho_s = 1$. При этом $\alpha = 0$, $p_{ош} = 0{,}5$. Этот случай соответствует одинаковым сигналам $s_1(t)$ и $s_2(t)$. Естественно, что различить такие сигналы невозможно. В результате приема таких сигналов апостериорная вероятность не увеличивается по сравнению с априорной и, следовательно, никакой полезной информации на выходе оптимального приемника нет.

2. $\rho_s = 0$. При этом $\alpha^2{}_а = E/N_0 = h^2{}_0$,

$$p_{ош} = 0{,}5 - \Phi(h_0). \qquad (3.104)$$

Этот случай соответствует классу двух ортогональных сигналов. Примерами таких сигналов являются частотноманипулированные (ЧМн), фазоманипулированные (ФМн) сигналы с манипуляцией на $\pi/2$ и др. [12].

3. $\rho_s = -1$. При этом $\alpha^2{}_а = 2E/N_0 = 2h^2{}_0$,

$$p_{ош} = 0{,}5 - \Phi(\sqrt{2}\, h_0). \qquad (3.105)$$

Этот случай соответствует классу противоположных сигналов $s_1(t) = -s_2(t)$, при которых достигается максимальное различие между s_1 и s_2. Примером таких сигналов являются фазоманипулированные сигналы с манипуляцией фазы на π.

При прочих равных условиях противоположные сигналы обеспечивают минимальную вероятность ошибки

по сравнению с любыми другими двоичными сигналами и в этом смысле они могут быть названы оптимальными двоичными сигналами.

Таким образом, рассмотрение вопросов оптимального приема двоичных сигналов позволило не только установить структуру оптимальных приемников, но и отыскать оптимальный класс двоичных сигналов. Следовательно, для двоичных сигналов полностью решена задача оптимизации системы типа «модулятор — демодулятор» (или, что эквивалентно, системы «сигнал — фильтр»).

Вероятность ошибки при оптимальном приеме двоичных сигналов разных классов часто удобно выражать в некотором модифицированном универсальном виде. Для этого аргумент α в формуле (3.99) представляют в виде

$$\alpha=\gamma_c h_0, \tag{3.106}$$

где γ_c — постоянный коэффициент, величина которого зависит от класса рассматриваемых двоичных сигналов. В частности, если пиковая мощность сигналов ограничена, то для сигналов с пассивной паузой $\gamma_c=1/\sqrt{2}$; для ортогональных сигналов с активной паузой $\gamma_c=1$; для противоположных $\gamma_c=\sqrt{2}$.

С учетом (3.106) формулы (3.99) и (3.103) принимают соответственно следующий вид:

$$p_{\text{ош}}=0{,}5-\Phi(\gamma_c h_0), \tag{3.107}$$

$$p_{\text{ош}}=\frac{1}{\sqrt{2\pi}\,\gamma_c h_0}\exp\left(-\frac{\gamma^2_c h^2_0}{2}\right) \text{ для } \gamma_c h_0>3. \tag{3.108}$$

Удобство такого представления состоит в том, что появляется возможность наглядно сравнивать помехоустойчивость оптимального приема для разных классов двоичных сигналов.

Зависимость $p_{\text{ош}}=f(h_0)$ для разных классов двоичных сигналов представлена на рис. 3.12. Величина h_0 для сигналов с активной паузой (как ортогональных, так и противоположных) характеризует энергетические затраты на передачу одной двоичной единицы информации с заданной вероятностью ошибки $p_{\text{ош}}$, так как для таких сигналов $E=E_0$ *.

* Напомним, что E_0 — энергия сигнала на входе приемника, соответствующая одной двоичной единице информации, а E — энергия одной посылки сигнала (см. § 1.5).

Сигналы с пассивной паузой можно рассматривать как частный случай ортогональных сигналов. В таких сигналах энергия затрачивается только при излучении посылки $s_1(t)$. Поэтому с учетом априорной вероятности $p(s_1)=0{,}5$ для обеспечения одинаковой вероятности ошибки энергия посылки $s_1(t)$ должна быть вдвое больше, чем при ортогональных сигналах с активной паузой, т. е. $E_п=2E_а$ и $h^2_{0п}=2h^2_{0а}$ (при $p_{ош\,п}=p_{ош\,а}$). При таком

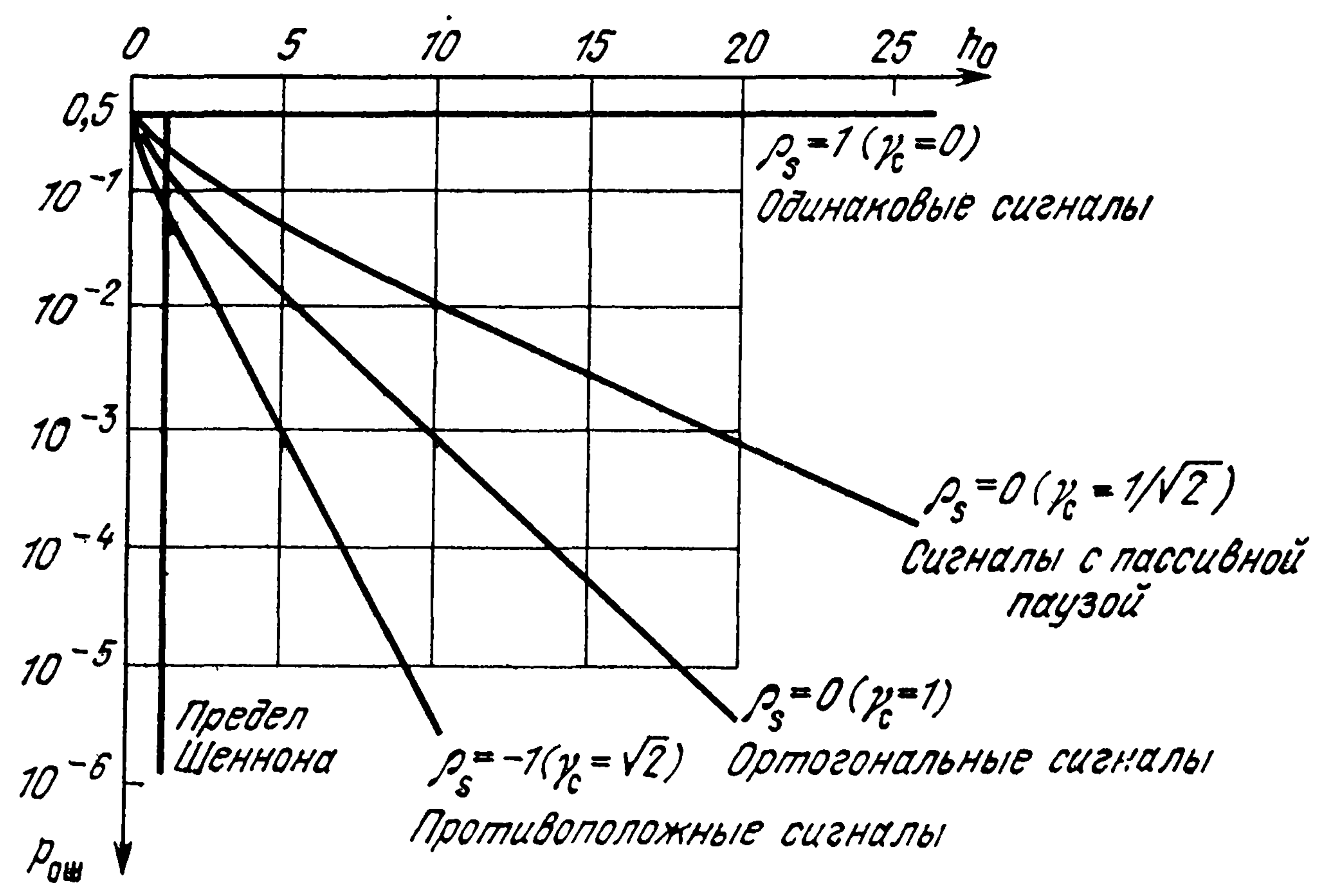


Рис. 3.12.

условии энергетические затраты на передачу одной двоичной единицы информации для ортогональных сигналов с активной и пассивной паузой одинаковы. Выполнить указанное условие можно, увеличивая либо среднюю (а следовательно, и пиковую) мощность посылки сигнала $s_1(t)$, либо длительность посылки τ_0. Ни то, ни другое во многих случаях нежелательно. Увеличение пиковой мощности связано с необходимостью повышения питающих напряжений в передатчике и усложнением его схемы, а увеличение длительности посылки τ_0 — с уменьшением скорости передачи информации. По этой и ряду других причин сигналы с пассивной паузой в современных системах цифровой передачи информации находят ограниченное применение.

Проведенное рассмотрение показывает, что помехоустойчивость приема при выбранном классе двоичных сигналов зависит только от величины энергетического отношения сигнала к шуму h^2_0.

Учтем, что энергия посылки

$$E=\int_0^{\tau_0} s^2(t)\,dt=\tau_0\frac{1}{\tau_0}\int_0^{\tau_0} s^2(t)\,dt=P_c\tau_0,$$

где P_c — средняя мощность сигнала на интервале времени $(0-\tau_0)$.

Обозначим полосу частот, занимаемую спектром сигнала, через $\Delta f_э$ (ширина спектра сигнала). Тогда h^2_0 можно представить в виде

$$h^2_0=\frac{E}{N_0}=\frac{P_c\tau_0}{N_0}=\frac{P_c}{N_0\Delta f_э}\Delta f_э\tau_0=\left(\frac{P_c}{P_ш}\right)_{вх}Б_c, \quad (3.109)$$

где

$$Б_c=\Delta f_э\tau_0 \quad (3.110)$$

— база сигнала, а $(P_c/P_ш)_{вх}$ — отношение средней мощности сигнала к средней мощности шума на входе приемника, взятое в полосе частот $\Delta f_э$.

Из выражения (3.109) следует, что необходимая величина h^2_0 может быть получена увеличением либо отношения сигнал/шум на входе оптимального приемника, либо базы сигнала.

Для сигналов с большой базой ($Б_c \gg 1$)

$$(P_c/P_ш)_{вх}=h^2_0/Б_c \ll h^2_0. \quad (3.111)$$

Для простых сигналов ($Б_c \approx 1$)

$$(P_c/P_ш)_{вх}=h^2_0. \quad (3.112)$$

Графики зависимости $p_{ош}=f(h_0)$, представленные на рис. 3.12, справедливы для любых двоичных сигналов, в том числе и простых и сложных. Эти графики характеризуют помехоустойчивость оптимального приема различных классов двоичных сигналов.

Зная вероятность ошибки при передаче одного символа дискретного сообщения, можно найти вероятность искажения кодовой комбинации первичного кода. Так как ошибки при передаче отдельных символов независимы, то при m-значном коде эта вероятность равна (при $p_{ош} \ll 1$)

$$p_{ошk}=1-(1-p_{ош})^m \approx mp_{ош}=m\,[0{,}5-\Phi(\gamma_c h_0)]. \quad (3.113)$$

На рис. 3.12 показана граница, определяющая предел Шеннона при передаче одной двоичной единицы инфор-

мации по гауссовскому каналу (см. § 1.7). Из рассмотрения графиков рис. 3.12 следует, что оптимальная система типа «модулятор — демодулятор» весьма существенно уступает полностью оптимальной системе в смысле Шеннона по величине энергетических затрат h^2_0 на передачу одной двоичной единицы информации. Так, например, для передачи с ошибкой от 10^{-4} до 10^{-6} необходимо иметь $h^2_0=7—11$, в то время как в полностью оптимальной системе при $p_{ош}=0$ достаточно иметь $h^2_0=$ $=\ln 2\approx 0{,}7$, т. е. в 10—15 раз меньше. Однако полная реализация подобного выигрыша принципиально невозможна, поскольку для этого необходимо применять кодирование, требующее бесконечного времени задержки. Вместе с тем, в настоящее время известны пути и способы, позволяющие достаточно близко подойти к предельным возможностям передачи информации. С некоторыми из них мы познакомимся в гл. 6.

3.6. КРАТКИЕ ИТОГИ ГЛ. 3.

1. В этой главе было подробно рассмотрено одно из наиболее простых приложений теории оптимального различения сигналов, а именно, оптимальное различение двух известных точно сигналов, передаваемых по каналу с постоянными параметрами, в котором единственной помехой является нормальный белый шум.

2. Проведенное рассмотрение позволило полностью решить (в рамках установленных допущений) задачу оптимизации цифровой системы связи без кодирования, т. е. определить структуру оптимальных приемников двоичных сигналов и найти соотношения, характеризующие качество работы таких приемников.

3. Рассмотрение качества работы оптимальных приемников, характеризуемого вероятностью ошибки при передаче одной двоичной единицы информации, дало возможность найти оптимальный класс двоичных сигналов. Такими сигналами являются противоположные сигналы. Эти сигналы обеспечивают при прочих равных условиях минимальную вероятность ошибки. Следовательно, установление оптимального класса двоичных сигналов, принимаемых на оптимальный приемник, равнозначно решению задачи оптимизации системы «модулятор — демодулятор» (или, что эквивалентно, системы «сигнал—фильтр») в классе двоичных сигналов.

4. Для любых классов двоичных сигналов оптимальный приемник может быть реализован на основе либо коррелятора, либо согласованного линейного фильтра. В первом случае вся фильтрация сигнала на фоне шума выполняется на видеочастоте, а во втором — на несущей частоте посылок.

5. Теория оптимального различения двух сигналов верна при существенных допущениях и ограничениях, которые практически никогда не могут быть выполнены точно. Несмотря на это, результаты теории имеют большую принципиальную важность по следующим причинам:

а) Проведенное рассмотрение позволяет в наиболее ясной форме продемонстрировать подход к решению задач оптимального приема дискретной информации. Идеи, изложенные в этой главе, лежат в основе исследования многих более сложных вопросов. К ним, в частности, относятся вопросы оптимального приема дискретной информации, представленной кодами с основанием больше двух (многоосновные коды). Оптимальные приемники в этом случае получаются обобщением рассмотрения, выполненного для двоичных сигналов *.

б) Результаты, полученные для сигналов с полностью известными параметрами, можно рассматривать как некоторые теоретические границы, с которыми удобно сравнивать более сложные случаи, когда сигналы содержат неизвестные мешающие параметры, обусловленные случайными изменениями характеристик канала и т. п.

в) Имеется ряд каналов передачи цифровой информации, которые с некоторым приближением можно считать каналами с постоянными параметрами. К таким каналам относятся каналы дальней и ближней космической связи; радиорелейные и кабельные каналы. Поэтому результаты оптимального приема сигналов, известных точно, можно рассматривать для таких каналов как достаточно близкую нижнюю границу.

КОНТРОЛЬНЫЕ ВОПРОСЫ И ЗАДАЧИ

3.1. Сформулируйте задачу различения двух сигналов на фоне помех.

3.2. Назовите основные допущения при решении классической задачи различения сигналов.

---

* Некоторые вопросы применения многоосновных кодов и соответствующих им сигналов рассмотрены в гл. 6.

3.3. В чем суть метода проверки гипотез?
3.4. Сформулируйте байесовское правило выброса гипотез. Почему такое правило обеспечивает минимум вероятности ошибки при выборе гипотез?
3.5. Что называют отношением правдоподобия и как его получить?
3.6. Докажите справедливость выражения (3.21).
3.7. Какие операции выполняет коррелятор оптимального приемника?
3.8. Покажите, что структура оптимального приемника не изменяется, если образец сигнала, подаваемого на коррелятор, усилить в K_0 раз (K_0=const).
3.9. Определите структуру оптимального приемника сигналов с активной паузой для следующих условий: а) $E_1 \neq E_2$,

$$p(s_1) = p(s_2) = 0{,}5; \quad б)\ E_1 = E_2,\ p(s_1) \neq p(s_2).$$

3.10. Получите формулы для вероятности ошибки в случае приема на оптимальные приемники, соответствующие условиям задачи 3.9.
3.11. Определите характеристики оптимального линейного фильтра для следующих видов посылок сигнала:

а) косинусоидальная огибающая

$$s(t) = \begin{cases} S_0 \cos(\pi t/\tau_0) \sin \omega_0 t, & -\tau_0/2 \leqslant t \leqslant \tau_0/2, \\ 0 & \text{для других } t; \end{cases}$$

б) прямоугольная огибающая

$$s(t) = \begin{cases} S_0 \sin \omega_0 t, & -\tau_0/2 \leqslant t \leqslant \tau_0/2, \\ 0 & \text{для других } t. \end{cases}$$

В обоих случаях принять, что величина $\omega_0\tau_0/2\pi = f_0\tau_0$ является целым числом.
3.12. Приведите примеры простых и сложных ортогональных сигналов с частотной и фазовой манипуляцией. Запишите выражения для таких сигналов.
3.13. Найдите выражение для коэффициента взаимной корреляции простых сигналов с частотной манипуляцией

$$s(t) = \begin{cases} s_1(t) = S_0 \sin(\omega_1 t + \varphi_1) \\ s_2(t) = S_0 \sin(\omega_2 t + \varphi_2) \end{cases}, \quad 0 \leqslant t \leqslant \tau_0.$$

Определите условия, при которых помехоустойчивость оптимального приемника таких сигналов достигает максимального значения.
3.14. Покажите, что коэффициент взаимной корреляции простых сигналов с фазовой манипуляцией

$$s(t) = \begin{cases} s_1(t) = S_0 \sin(\omega_0 t + \varphi_1) \\ s_2(t) = S_0 \sin(\omega_0 t + \varphi_2) \end{cases}, \quad 0 \leqslant t \leqslant \tau_0,$$

определяется формулой

$$\rho_в = \cos(\varphi_1 - \varphi_2) = \cos \Delta\varphi.$$

3.15. Покажите, что вероятность ошибки оптимального приема ФМн сигналов, рассмотренных в предыдущей задаче, при условии $h^2_0 \geqslant 9/(1-\rho_s)$ можно определить по формуле

$$p_{\text{ош}} \approx 0,5 \exp\left(-h^2_0 \sin^2 \frac{\Delta\varphi}{2}\right)\left(\sqrt{\pi} h_0 \sin \frac{\Delta\varphi}{2}\right)^{-1}.$$

3.16. Получите приближенную формулу для вероятности ошибки оптимального приема простых ЧМн сигналов (см. задачу 3.13) при условии, что $h_0 \geqslant 3$.

3.17. Какой вид имеют зависимости $p_{\text{ош}}=f(\gamma_c h_0)$ и $p_{\text{ош}}=f(\gamma^2{}_c h^2{}_0)$, если вероятность ошибки определять логарифмической мерой?

СПИСОК ЛИТЕРАТУРЫ

1. **Котельников В. А.** Теория потенциальной помехоустойчивости. М., Госэнергоиздат, 1956.
2. **Миддлтон Д.** Введение в статистическую теорию связи. Т. 2. М., «Сов. радио», 1962.
3. **Тихонов В. И.** Статистическая радиотехника. М., «Сов. радио», 1966.
4. **Гуткин Л. С.** Теория оптимальных методов радиоприема при флюктуационных помехах. Изд. второе. М., «Сов. радио», 1972.
5. **Ван Трис Г.** Теория обнаружения, оценок и модуляции. Т. 1. М., «Сов. радио», 1972.
6. **Давенпорт В., Рут В.** Введение в теорию случайных сигналов и шумов. М., ИЛ, 1960.
7. **Левин Б. Р.** Теоретические основы статистической радиотехники. Т. 2. М., «Сов. радио», 1968.
8. **Витерби Э.** Принципы когерентной связи. М., «Сов. радио», 1970.
9. **Вентцель Е. С.** Теория вероятностей. М., «Наука», 1969.
10. **Корн Г., Корн Т.** Справочник по математике для научных работников и инженеров. М., «Наука», 1970.
11. **Абезгауз Г. Г., Тронь А. П., Копенкин Ю. Н., Коровина И. А.** Справочник по вероятностным расчетам. М., Воениздат, 1970.
12. **Финк Л. М.** Теория передачи дискретных сообщений. М., «Сов. радио», 1970.

Глава 4

Реальные способы приема простых двоичных сигналов в каналах с постоянными параметрами

4.1. ВВЕДЕНИЕ

В предыдущей главе были подробно рассмотрены особенности оптимального приема двоичных сигналов в каналах с постоянными параметрами. Несмотря на простоту структуры оптимальных приемников их реализация оказывается во многих случаях сложной технической проблемой. Это объясняется тем, что для построения оптимального приемника необходимо точно знать несущую частоту принимаемого сигнала и его начальную фазу, а также быть уверенным в неизменности амплитуды принимаемых сигналов на интервале анализа. Строго говоря, реальных каналов с постоянными параметрами не существует. В реальных каналах параметры принимаемых сигналов всегда приобретают некоторые случайные отклонения, обусловленные нестабильностью характеристик передающего устройства, механическими колебаниями антенных устройств, изменениями параметров среды распространения, перемещениями передатчика или приемника в пространстве (при связи с подвижными объектами) и т. д. Такие отклонения приводят к «рассогласованию» между параметрами принятого сигнала и характеристиками оптимального приемника, построенного в соответствии с заранее известными «образцами» посылок сигнала *.

Оптимальные приемники весьма чувствительны к такому «рассогласованию», и, если оно превышает некоторое значение, их показатели резко ухудшаются. В зависимости от структуры оптимального приемника ука-

* Здесь речь идет о сравнительно небольших случайных отклонениях параметров от их номинальных значений. В иностранной литературе такие отклонения часто называют «джиттер» (jitter — дрожание, пульсация). По сути дела в этой главе рассматриваются такие каналы, которые лишь в первом приближении можно считать каналами с постоянными параметрами. Типичные каналы со случайными параметрами рассматриваются в гл. 5.

занное рассогласование проявляет себя по-разному. Приемники с согласованными фильтрами очень критичны к небольшим рассогласованиям между моментом взятия отсчета выходного напряжения и моментом, когда это напряжение достигает максимума. Сказанное поясняется рис. 4.1. При незначительном отклонении момента взятия отсчета от значения $t_k=\tau_0$ величина

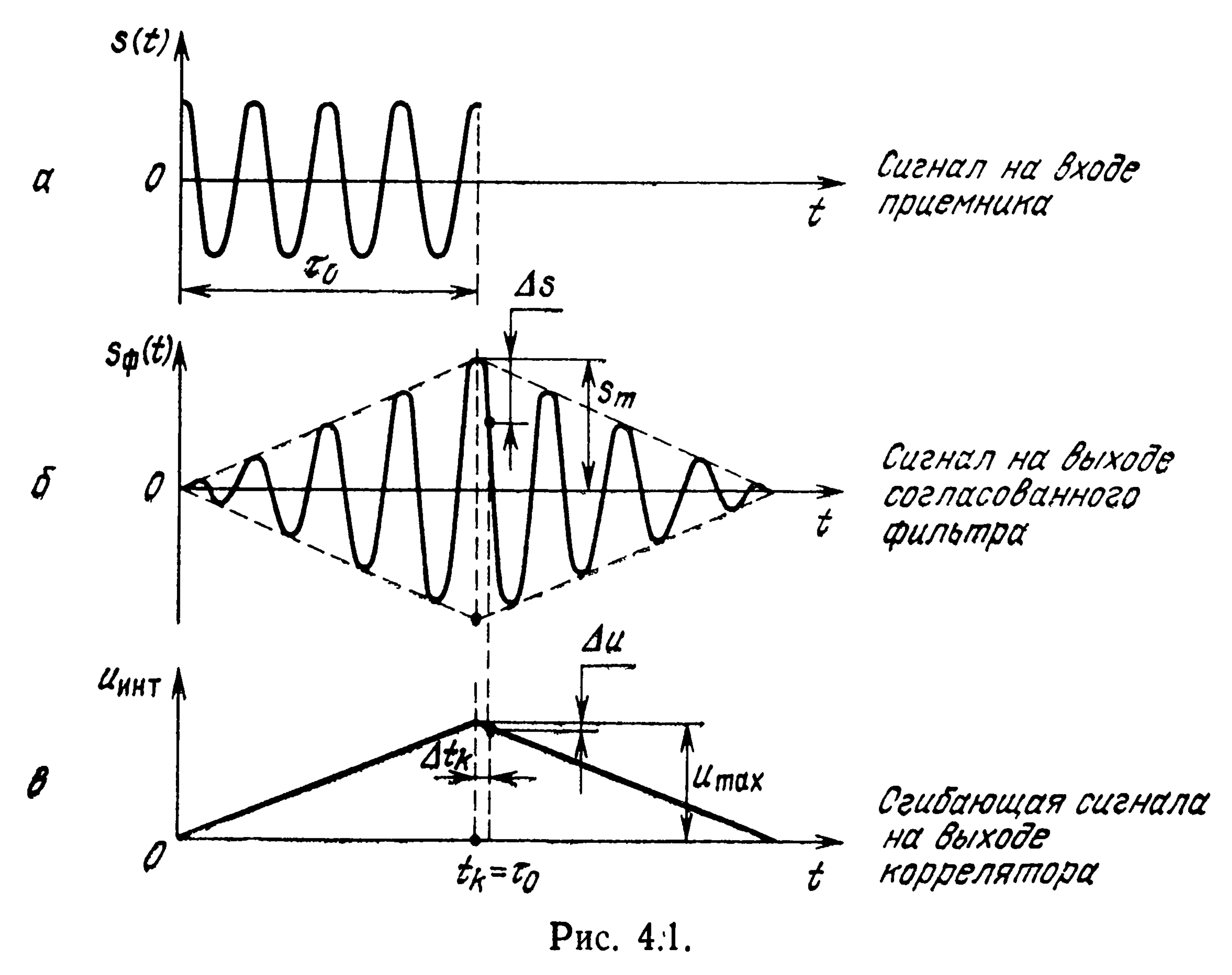


Рис. 4.1.

отсчета может резко уменьшиться по сравнению с максимальным значением и даже изменить свой знак (рис. 4.1,*б*). Вместе с тем, согласованные фильтры менее критичны к небольшим рассогласованиям по частоте и фазе.

Приемники с корреляторами наоборот менее критичны к нестабильности момента взятия отсчета (рис. 4.1,*в*), но весьма критичны к рассогласованию между параметрами принимаемого сигнала и параметрами «образцов» посылок сигнала, формируемых в корреляционном приемнике.

Для уменьшения влияния указанных видов рассогласования при реализации оптимальных приемников необходимо непрерывно оценивать (измерять) параметры принимаемого сигнала и устранять появляющееся рассогласование. Эта процедура связана с заметным услож-

нением приемного устройства и увеличением его стоимости, что в ряде случаев не оправдывается тем повышением помехоустойчивости, которое достигается по сравнению с более простым приемником. Поэтому при реализации приемных устройств в каналах с незначительными изменениями параметров часто и не стремятся к тому, чтобы полностью сохранить структуру приемников, обеспечивающих оптимальный прием с учетом этих изменений *.

Комбинируя положительные свойства оптимальных приемников с согласованными фильтрами и корреляторами, можно создать приемное устройство, в котором основную фильтрацию сигнала от помехи выполняет квазиоптимальный фильтр, а отсчет в конце интервала анализа входного колебания $y(t)$ берется по огибающей посылки.

Как известно, квазиоптимальными называют такие линейные фильтры, частотная и фазовые характеристики которых заданы или выбраны заранее, и максимум отношения сигнала к шуму обеспечивается только выбором полосы пропускания. Исследования показывают, что для простых двоичных сигналов ** квазиоптимальные фильтры дают почти такие же результаты, как и оптимальные (проигрыш не превышает 1 дБ) [1, 5]. В то же время квазиоптимальные фильтры малокритичны к сравнительно небольшим отклонениям частоты и некритичны к отклонениям фазы принимаемого сигнала.

Для получения огибающей посылки сигнала можно использовать обычный детектор огибающей или синхронный (когерентный) детектор. В последнем случае

---

* Заметим, что можно построить оптимальный приемник, отказавшись от оценки случайных изменений параметров сигнала. Например, для сигнала со случайной начальной фазой и т. п. Естественно, показатели качества работы такого приемника окажутся ниже показателей оптимального приемника, учитывающего эти изменения. Некоторые вопросы одиночного оптимального приема в каналах со случайными параметрами рассмотрены в гл. 5. Здесь мы на них останавливаться не будем.

** Простыми двоичными сигналами называют сигналы, посылки которых на интервале $[0, \tau_0]$ представляют собой синусоидальные колебания с параметрами, принимающими только одно из двух возможных значений (например, частота ω_1 или ω_2; амплитуда $S=S_0$ или $S=0$; фаза φ_1 или φ_2). Для таких сигналов $\Delta f_{опт}\tau_0 \approx 1$, где $\Delta f_{опт}$ — оптимальная полоса пропускания фильтра для посылки длительностью τ_0.

необходимо иметь опорное напряжение, аналогичное «образцу» сигнала в оптимальном корреляционном приемнике.

Основываясь на изложенных соображениях, иногда удается значительно упростить реализацию приемников (а, следовательно, уменьшить их стоимость), сделать эти приемники малочувствительными к случайным изменениям некоторых характеристик канала связи и параметров принимаемого сигнала и в то же время не очень ухудшить их качественные показатели по сравнению с соответствующими оптимальными приемниками.

Основное внимание в этой главе будет уделено некоторым наиболее применяемым в практике способам приема простых двоичных сигналов. Ниже такие способы будем называть реальными способами приема. Структура приемных устройств при этом задается заранее на основе учета ряда факторов инженерного характера, о которых шла речь выше.

Существующие реальные способы приема можно разделить на две группы: когерентный прием и некогерентный прием. Эти виды приема могут быть применены для любых видов манипулированных сигналов. Однако для простых двоичных АМн и ЧМн сигналов когерентные способы приема обычно не используются. Это объясняется тем, что схемы когерентного приема значительно сложнее схем некогерентного, и если уж идти на такое усложнение, то целесообразнее применять ФМн сигналы. Поэтому в данной главе для простых АМн и ЧМн сигналов рассмотрены только некогерентные способы приема, а для ФМн сигналов — когерентный способ приема и его разновидности.

4.2. НЕКОГЕРЕНТНЫЙ ПРИЕМ ПРОСТЫХ АМПЛИТУДНО-МАНИПУЛИРОВАННЫХ СИГНАЛОВ

При передаче двоичной цифровой информации простыми АМн сигналами на вход приемника поступает колебание

$$y(t)=s_i(t)+n(t), \qquad (4.1)$$

где сигнал $s_i(t)$ на интервале времени $0\leqslant t\leqslant\tau_0$ можно определить выражением

$$s_i(t)=S_0x_i\sin(\omega_0t+\varphi_0), \qquad (4.2)$$

которое принимает значения

$$s_i(t)=\begin{cases} s_1(t)=S_0\sin(\omega_0 t+\varphi_0), & \text{если } x_i=x_1=1, \\ s_2(t)\equiv 0, & \text{если } x_i=x_2=0. \end{cases} \qquad (4.3)$$

Структурная схема приемника, осуществляющего некогерентный прием простых АМн сигналов, имеет вид, показанный на рис. 4.2. Додетекторную часть приемника, состоящую в общем случае из входных цепей, усили-

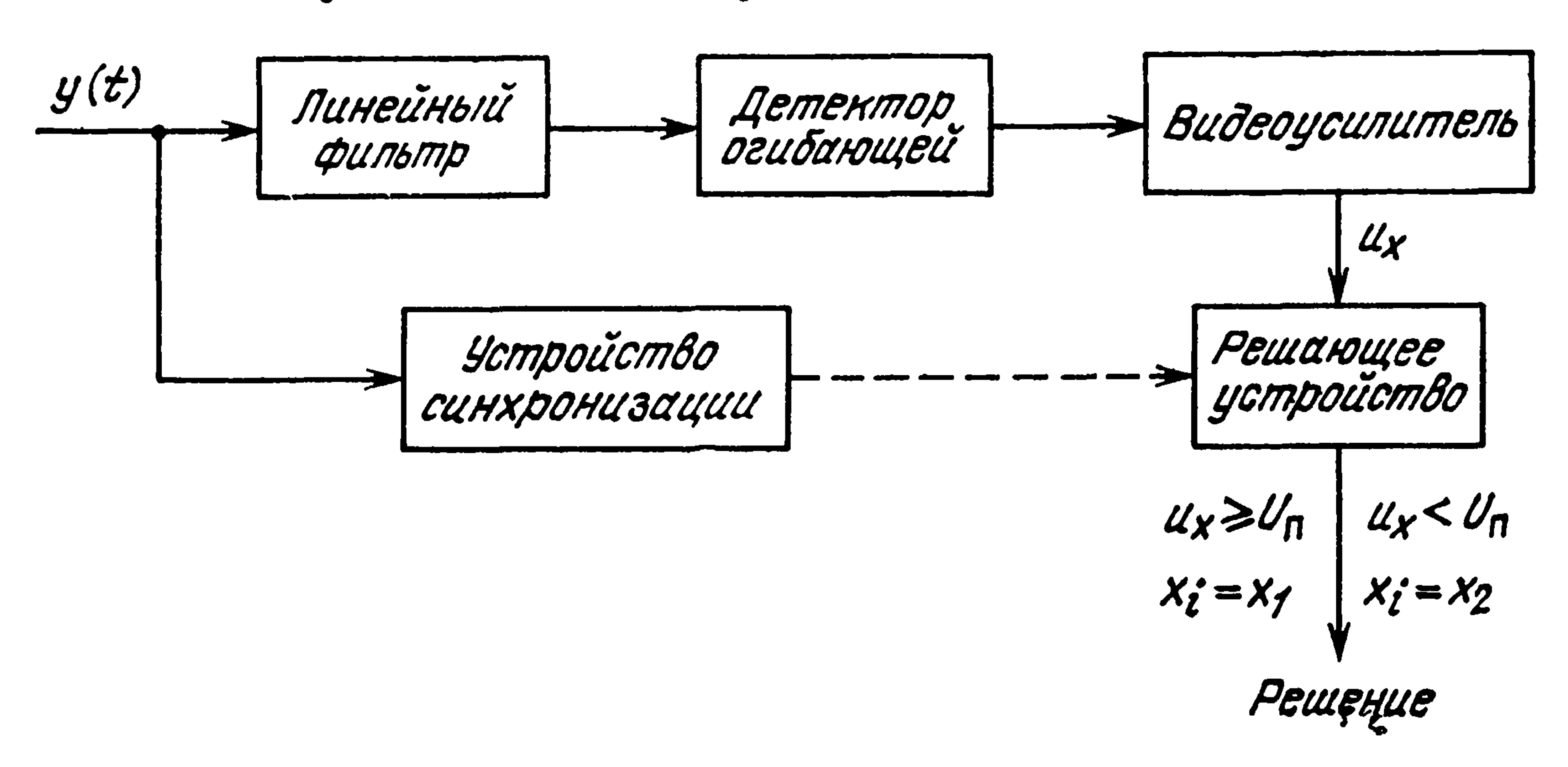


Рис. 4.2.

теля высокой частоты, преобразователя и усилителя промежуточной частоты (УПЧ), можно рассматривать как линейный фильтр, характеристики которого в основном определяются частотной и фазовой характеристиками УПЧ. Эта часть приемника обеспечивает основную селекцию сигнала от помех. Видеоусилитель обеспечивает фильтрацию огибающей посылок сигнала от высокочастотных составляющих помехи и продуктов детектирования на выходе детектора, а также необходимое усиление огибающей. С выхода видеоусилителя случайное напряжение u_x поступает на решающее устройство.

При некогерентном приеме АМн сигналов их различение осуществляется сравнением величины u_x, выделенной детектором, с некоторым пороговым напряжением $U_п$ и принятием решения по правилу

$$u_x \gtrless U_п \quad \begin{matrix} s_i(t)=s_1(t), & x_i=x_1, \\ s_i(t)=s_2(t), & x_i=x_2. \end{matrix} \qquad (4.4)$$

В соответствии с этим правилом решение о том, что был передан сигнал s_1, принимается при $u_x \geqslant U_п$, а сигнал s_2 — при $u_x < U_п$.

Так же, как и прежде, полагаем, что априорные вероятности посылок s_1 и s_2 одинаковы, т. е. $p(s_1)=$ $=p(s_2)=0{,}5$. Тогда вероятность ошибки при передаче одного символа сообщения определяется выражением (3.51).

Обозначим огибающую u_x при передаче сигнала s_1 через $U_{сш}$, а при передаче сигнала $s_2\equiv 0$ — через $U_ш$, т. е.

$$u_x=\begin{cases}U_{сш} \text{ для } y(t)=s_1(t)+n(t),\\ U_ш \text{ для } y(t)=n(t).\end{cases}$$

Тогда условные вероятности ошибок можно записать в виде

$$p(s_2|s_1)=p(U_{сш}\leqslant U_п)=\int\limits_0^{U_п} W_1(U_{сш})\,dU_{сш}, \qquad (4.5)$$

$$p(s_1|s_2)=p(U_ш\geqslant U_п)=\int\limits_{U_п}^{\infty} W_1(U_ш)\,dU_ш, \qquad (4.6)$$

где $W_1(U_{сш})$ и $W_1(U_ш)$ — одномерные распределения плотности вероятностей огибающих $U_{сш}$ и $U_ш$ соответственно.

Как известно [3—5], огибающая распределения смеси синусоидального колебания с амплитудой S_0 и нормального шума подчиняется распределению Райса:

$$W(U_{сш})=\frac{U_{сш}}{\sigma^2}\exp\left(-\frac{U^2_{сш}+S^2_0}{2\sigma^2}\right)I_0\left(\frac{U_{сш}S_0}{\sigma^2}\right), \qquad (4.7)$$

где σ^2 — средний квадрат шума; I_0 — функция Бесселя нулевого порядка от мнимого аргумента.

Распределение (4.7) часто также называют обобщенным распределением Релея. Если $S_0\to 0$ $(I_0\to 1;$ $U_{сш}\to U_ш)$, распределение (4.7) принимает вид

$$W_1(U_ш)=\frac{U_ш}{\sigma^2}\exp\left(-\frac{U^2_ш}{2\sigma^2}\right). \qquad (4.8)$$

Это распределение Релея. Оно описывает плотность вероятностей огибающей нормального шума.

Введем в (4.7) и (4.8) обозначения

$$U_{сш}/\sigma=v,\ S_0/\sigma=a,\ U_ш/\sigma=z. \qquad (4.9)$$

Тогда выражения (4.7) и (4.8) примут вид

$$W_1(v) = v \exp\left(-\frac{v^2 + a^2}{2}\right) I_0(av), \qquad (4.10)$$

$$W_1(z) = z \exp(-z^2/2). \qquad (4.11)$$

Вид этих распределений показан на рис. 4.3. При увеличении энергии сигнала s_1 кривая для $W_1(v)$ смещается вправо (пунктирная кривая).

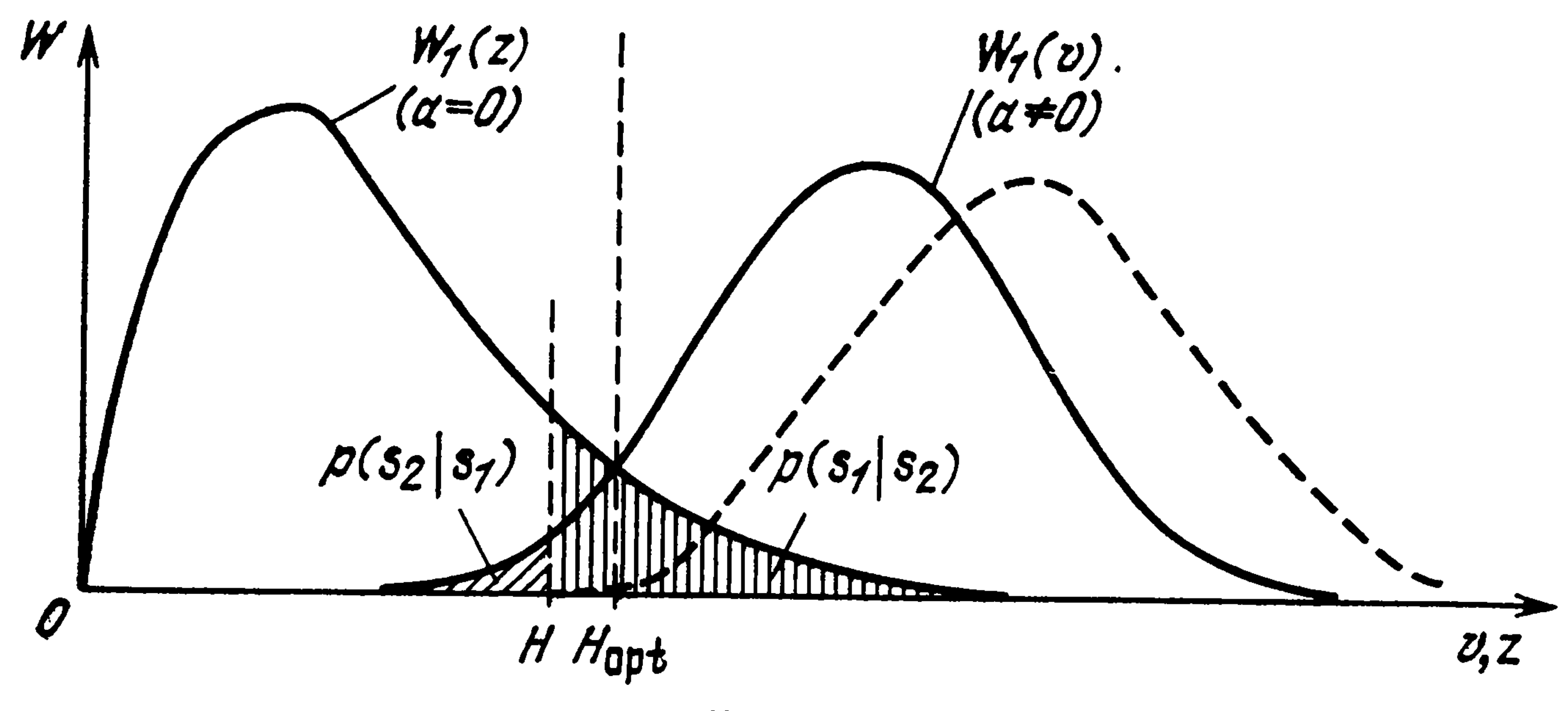


Рис. 4.3.

Приняв во внимание (4.5) и (4.6), получим

$$p(s_2|s_1) = \int_0^H \exp\left(-\frac{v^2 + a^2}{2}\right) I_0(av)\, dv, \qquad (4.12)$$

$$p(s_1|s_2) = \int_H^\infty z \exp\left(-\frac{z^2}{2}\right) dz = \exp\left(-\frac{H^2}{2}\right). \qquad (4.13)$$

Здесь величина Н — относительный порог, равный

$$H = U_п/\sigma. \qquad (4.14)$$

Выражение (4.12) часто называют интегральным распределением Райса. При фиксированном значении порога H вероятности $p(s_2|s_1)$ и $p(s_1|s_2)$ определяются площадью заштрихованных участков на рис. 4.3.

Интеграл (4.12) для конечных значений H не берется, поэтому в литературе по теории статистического приема для представления подобных интегралов широко пользуются табулированной функцией, которую называют Q-функцией [6, 7]. Под Q-функцией понимают

дополнение интегрального распределения Райса до единицы. В нашем случае

$$Q(a, H) = \int_{H}^{\infty} v \exp\left(-\frac{v^2 + a^2}{2}\right) I_0(av)\, dv. \quad (4.15)$$

Нетрудно видеть, что Q-функция обладает следующими свойствами:

$$\begin{aligned} &1)\ Q(a, H) \to Q(a, 0) = 1 \quad \text{при } H \to 0, \\ &2)\ Q(a, H) \to Q(0, H) = \exp(-H^2/2) \quad \text{при } a \to 0. \end{aligned} \quad (4.16)$$

Введенная функция позволяет записать выражение (4.12) в виде

$$p(s_2|s_1) = 1 - Q(a, H). \quad (4.17)$$

Рассмотрение выражений (4.12) и (4.13) показывает, что условные вероятности ошибок $p(s_2|s_1)$ и $p(s_1|s_2)$ различаются между собой. Это значит, что при некогерентном приеме АМн сигналов канал несимметричен.

С учетом (4.13) и (4.17) выражение для вероятности ошибки при некогерентном приеме простых АМн сигналов принимает вид

$$p_{\text{ош}} = 0{,}5[1 - Q(a, H) + \exp(-H^2/2)]. \quad (4.18)$$

Проанализируем это выражение. Сигналы с амплитудной манипуляцией являются сигналами с пассивной паузой ($s_2 \equiv 0$). Для таких сигналов вероятность $p(s_2|s_1)$ характеризует вероятность пропуска сигнала s_1, а $p(s_1|s_2)$ — вероятность ложного появления сигнала s_1 (ложная тревога). Нетрудно видеть, что для предельных значений порога H имеем

1) $p(s_1|s_2) \longrightarrow 1$, $p(s_2|s_1) \longrightarrow 0$, $p_{\text{ош}} \longrightarrow 0{,}5$ при $H \longrightarrow 0$;
2) $p(s_1|s_2) \longrightarrow 0$, $p(s_2|s_1) \longrightarrow 1$, $p_{\text{ош}} \longrightarrow 0{,}5$ при $H \longrightarrow \infty$.

Таким образом, при очень низких значениях порога ($H \longrightarrow 0$) основную долю ошибок создают ложные появления сигнала s_1, а при очень высоких ($H \longrightarrow \infty$) — пропуски сигнала s_1. В первом случае на выходе приемника формируется последовательность символов, состоящая только из единиц, а во втором — только из нулей. Иными словами, в первом случае приемное устройство полностью «открыто» для любого воздействия, независимо от того определяется ли оно только шумом или смесью сигнала с шумом. Во втором случае приемник полно-

стью «закрыт» и не реагирует ни на какое воздействие. Очевидно, что и в том, и в другом случае независимо от величины отношения сигнал/шум на входе приемника полезная информация на выходе приемника полностью отсутствует*. Для промежуточных значений порога ($0<H<\infty$) вероятность ошибки зависит как от значения порога H, так и от отношения сигнал/шум на входе приемника.

Учитывая рассмотренные выше особенности работы приемника, нетрудно убедиться в том, что для каждого фиксированного отношения сигнал/шум должно существовать оптимальное значение порога, при котором вероятность ошибки принимает минимальное значение. Выражение для оптимального порога H можно найти, исследуя функцию (4.18) на минимум. Однако задачу можно решить проще. Из рассмотрения рис. 4.3 следует, что минимальное значение вероятности ошибки получается, если порог H соответствует пересечению кривых распределений $W_1(v)$ и $W_1(z)$. Отклонение от этой точки пересечения в любую сторону увеличивает одну из условных вероятностей ошибок больше, чем уменьшает другую.

Выполнив условие

$$W_1(z)\Big|_{z=H_{\text{опт}}} = W_1(v)\Big|_{v=H_{\text{опт}}},$$

нетрудно получить уравнение для определения оптимального значения порога

$$I_0(aH_{\text{опт}}) = \exp(a^2/2). \tag{4.19}$$

Так как уравнение (4.19) является трансцендентным, то решить его в явном виде можно либо графически, либо численно.

Введем величину, характеризующую отношение мощности сигнала к мощности шума на входе приемника:

$$h^2 = (P_{\text{с}}/P_{\text{ш}})_{\text{вх}} = S_0^2/2\sigma^2 = S_0^2/2N_0\Delta f_{\text{э}}. \tag{4.20}$$

Сравнив выражения (4.20) и (4.9), получим

$$h = a/\sqrt{2}; \quad a = \sqrt{2}h. \tag{4.21}$$

Представим относительный порог в виде

$$H_{\text{опт}} = \frac{U_{\text{п опт}}}{\sigma} = \frac{U_{\text{п опт}}}{S_0} \cdot \frac{S_0}{\sigma} = \frac{U_{\text{п опт}}}{S_0} \sqrt{2}h. \tag{4.22}$$

---

* Напомним, что в бинарных каналах с одинаковыми априорными вероятностями посылок сигналов скорость передачи информации равна нулю, если $p_{\text{ош}}=0{,}5$.

С учетом записанных соотношений выражение (4.19) примет вид

$$I_0\left(2h^2\frac{U_{\text{п опт}}}{S_0}\right)=\exp(h^2). \tag{4.23}$$

На рис. 4.4 показана зависимость $U_{\text{п опт}}/S_0=f(h)$, рассчитанная в соответствии с выражением (4.23). При увеличении h кривая асимптотически стремится к значению 0,5.

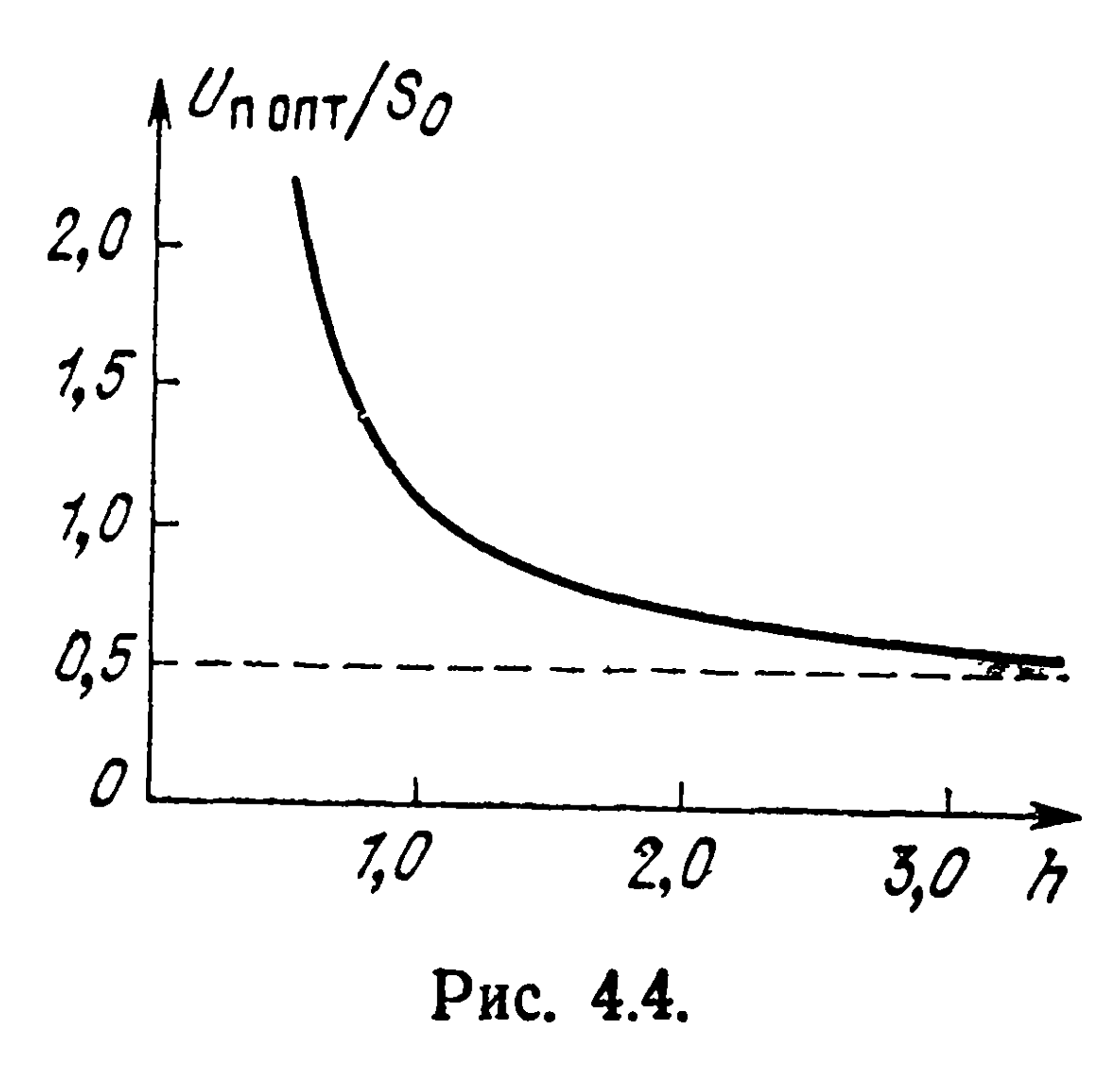


Рис. 4.4.

Как известно [8], при сравнительно небольшом отношении сигнал/шум в детекторе огибающей начинает заметно сказываться явление подавления сигнала шумом. Чтобы это явление практически не сказывалось, необходимо обеспечить условие

$$h\geqslant 3. \tag{4.24}$$

При таком условии

$$U_{\text{п опт}}/S_0\approx 0{,}5. \tag{4.25}$$

Следовательно, для значений $h\geqslant 3$

$$H_{\text{опт}}\approx h/\sqrt{2}=a/2. \tag{4.26}$$

При выполнении условия (4.24) $a>4$, и распределение Райса (4.10) становится близким к нормальному [3, 4], т. е. принимает вид

$$W_1(v)=\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{(v-a)^2}{2}\right]. \tag{4.27}$$

В этом случае

$$p(s_2|s_1)=1-Q(a,H)=1-\frac{1}{\sqrt{2\pi}}\int\limits_H^\infty \exp\left[-\frac{(v-a)^2}{2}\right]dv.$$

Заменив переменную интегрирования $v-a=y$, получим

$$p(s_2|s_1)=1-\frac{1}{\sqrt{2\pi}}\int\limits_{H-a}^\infty \exp\left(-\frac{y^2}{2}\right)dy.$$

При оптимальном пороге $(H=H_{\text{опт}})$

$$H_{\text{опт}}-a=-S_0/2\sigma=-H_{\text{опт}}.$$

Следовательно,

$$p(s_2|s_1) = 1 - \frac{1}{\sqrt{2\pi}} \int\limits_{-H_{\text{опт}}}^{\infty} \exp\left(-\frac{y^2}{2}\right) dy = 0,5 - \Phi(H_{\text{опт}}).$$

Здесь учтено, что

$$\frac{1}{\sqrt{2\pi}} \int\limits_{-H_{\text{опт}}}^{\infty} \exp\left(-\frac{y^2}{2}\right) dy = 0,5 + \Phi(H_{\text{опт}}).$$

Итак, вероятность ошибки при некогерентном приеме АМн сигналов при большом отношении сигнал/шум на входе приемника ($h>3$) и оптимальном пороге различения сигналов, определяется выражением

$$p_{\text{ош}} = 0,5[0,5 - \Phi(H_{\text{опт}}) + \exp(-H^2_{\text{опт}}/2)], \tag{4.28}$$

или с учетом (4.26)

$$p_{\text{ош}} = 0,5[0,5 - \Phi(h/\sqrt{2}) + \exp(-h^2/2)]. \tag{4.29}$$

Для значений $h>4$ последнее выражение можно упростить (см. формулу (3.102)):

$$p_{\text{ош}} \approx 0,5 \exp(-h^2/4)(1 + 1/\sqrt{\pi}\, h). \tag{4.30}$$

Чтобы обеспечить работу приемника АМн сигналов в режиме оптимального порога, необходимо знать величину отношения сигнал/шум на входе приемника и устанавливать соответствующее значение порога. Если величина h может изменяться в процессе передачи информации, например, при взаимном движении объектов, на которых размещены передатчик и приемник данной линии связи (динамический канал связи), то приемник должен иметь устройство оценки этой величины. Практически такая оценка выполняется автоматической регулировкой усиления (АРУ) приемника по принимаемому сигналу. Если мощность сигнала возрастает, то схема АРУ уменьшает усиление приемника так, чтобы при фиксированном значении напряжения порога $U_{\text{п}}$ вероятность тревог $p(s_1|s_2)$ уменьшалась. Это эквивалентно увеличению относительного порога H в соответствии с увеличением отношения сигнал/шум на входе приемника, т. е. сохранению условия оптимальности

порога. Нетрудно убедиться в том, что при уменьшении мощности сигнала схема АРУ также обеспечивает поддержание оптимального режима работы приемника.

Рассмотрение выражений (4.29) и (4.30) показывает, что для больших отношений сигнал/шум ($h>4$) вероятность ошибки при некогерентном приеме АМн сигналов приемником с оптимальным порогом в основном определяется вероятностью ложных появлений сигнала s_1 (ложной тревоги) и равна

$$p_{\text{ош}} \approx 0{,}5 \exp(-h^2/4), \quad h>3-4. \qquad (4.31)$$

Выясним, насколько проигрывает некогерентный прием АМн простых сигналов по сравнению с оптимальным приемом. Вероятность ошибки при оптимальном приеме АМн сигналов определяется выражением (3.80). Сравним эту вероятность с вероятностью ошибки, определяемой выражением (4.31). При сравнении будем полагать, что полоса пропускания некогерентного приемника до детектора выбрана оптимально, т. е. соответствует условию $\Delta f_э \tau_0 = 1$. Тогда

$$h^2 = (P_c/P_ш)_{вх} = S^2{}_0 \tau_0 / 2N_0 \Delta f_э \tau_0 = E/N_0 = h^2{}_0. \qquad (4.32)$$

Для удобства сравнения величину h_0 при некогерентном приеме обозначим $h^2{}_{0р}$.

Используя соотношения (3.80), (4.31) и (4.32), вероятности ошибок как при оптимальном, так и при некогерентном приеме для $h_0>4$ с достаточно хорошей точностью можно записать в виде

$$p_{\text{ош опт}} = \frac{1}{\sqrt{\pi}\, h_0} \exp(-h^2{}_0/4), \qquad (4.33)$$

$$p_{\text{ош нк}} \approx 0{,}5 \exp(-h^2{}_{0р}/4). \qquad (4.34)$$

Сравнение способов приема можно провести при разных условиях.

1. $h_{0р} = h_0$. Это условие означает одинаковые средние энергетические затраты при передаче одного элемента двоичной последовательности. В этом случае сравнение ведется по величине отношения вероятностей ошибок

$$\gamma_р = \frac{p_{\text{ош нк}}}{p_{\text{ош опт}}} = \frac{\sqrt{\pi}}{2} h_0. \qquad (4.35)$$

2. $p_{\text{ош нк}} = p_{\text{ош опт}}$. Сравнение при таком условии сводится к определению отношения энергетических затрат, обеспечивающих одинаковую вероятность ошибки при оптимальном и некогерентном способах приема.

Приравняв выражение (4.33) к (4.34), после логарифмирования и несложных преобразований можно получить

$$\gamma_h = \frac{h^2{}_{0p}}{h^2{}_0} = 1 + \frac{4}{h^2{}_0} \ln\left(\frac{\sqrt{\pi}}{2} h_0\right). \qquad (4.36)$$

Эта формула определяет энергетический проигрыш некогерентного способа приема по сравнению с оптимальным. Расчеты показывают, что для обеспечения вероятности ошибки от 10^{-3} до 10^{-6} некогерентный способ приема АМн сигналов требует увеличения энергии сигнала на (30—15)%. При большом отношении сигнал/шум разница между оптимальным и некогерентным приемом незначительна.

4.3. НЕКОГЕРЕНТНЫЙ ПРИЕМ ПРОСТЫХ ЧАСТОТНО-МАНИПУЛИРОВАННЫХ СИГНАЛОВ

Простые частотно-манипулированные сигналы, используемые для передачи двоичной цифровой информации, можно записать в виде

$$s(t) = \begin{cases} s_1(t) = S_0 \cos(\omega_1 t + \varphi_1) \\ s_2(t) = S_0 \cos(\omega_2 t + \varphi_2) \end{cases} \quad \text{при } 0 \leqslant t \leqslant \tau_0, \qquad (4.37)$$

где ω_1 и φ_1 — частота и фаза посылки, соответствующей передаче информационного символа 1; ω_2 и φ_2 — частота и фаза посылки, соответствующей символу 0.

Спектр простого ЧМн сигнала показан на рис. 4.5, где Δf_p — интервал разноса посылок s_1 и s_2 по частоте (девиация частоты). Можно показать, что имеется оптимальная величина разноса частот, при которой достигается максимальная помехоустойчивость оптимального приема ЧМн сигналов [9]*:

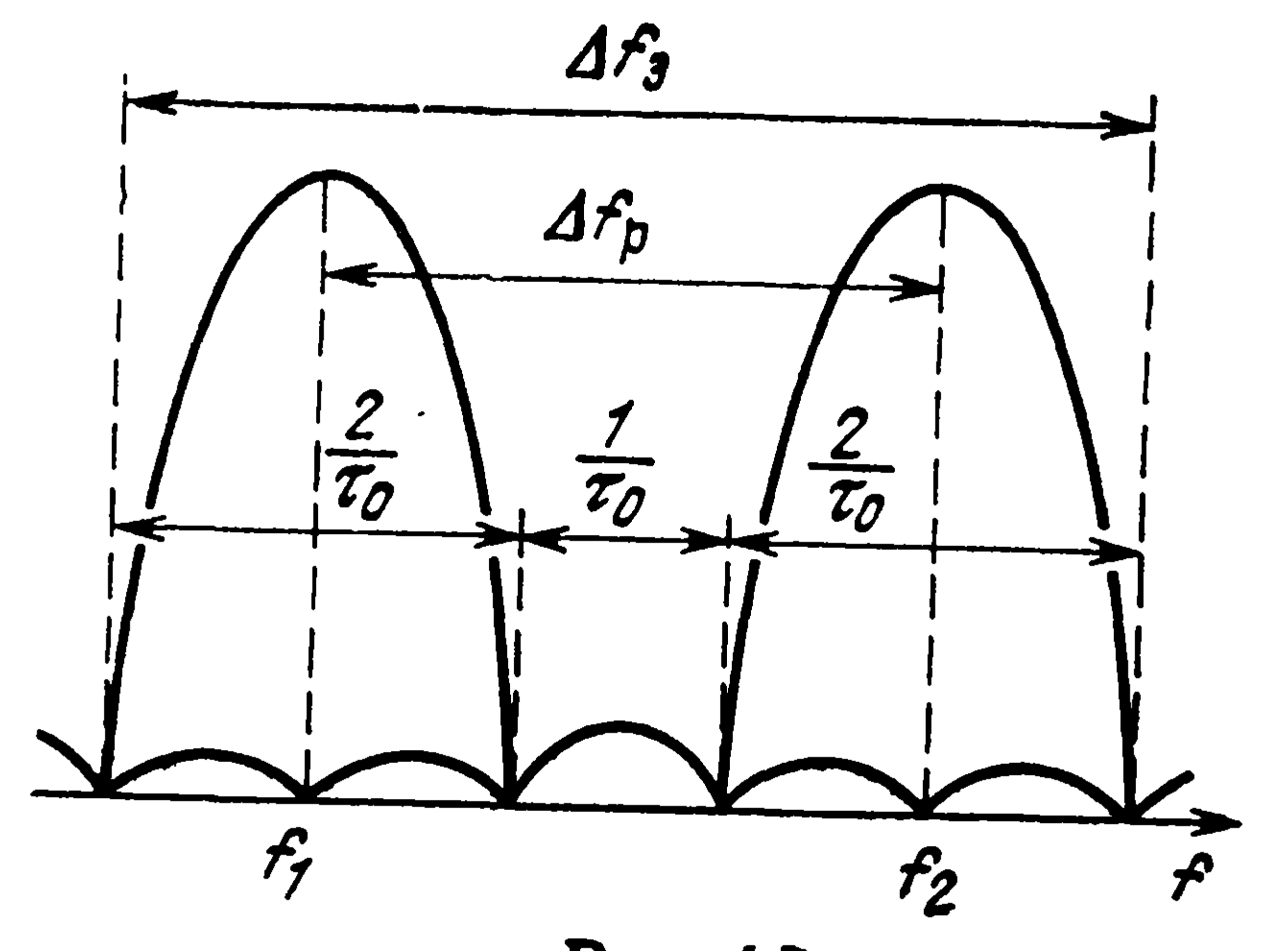


Рис. 4.5.

$$\text{opt}\,\Delta f_p = 0{,}75/\tau_0. \qquad (4.38)$$

* См. также задачу 3.13 в гл. 3.

Если $\Delta f_р < \text{opt}\,\Delta f_р$, то условия различения сигналов s_1 и s_2 заметно ухудшаются из-за значительного перекрытия спектров посылок s_1 и s_2 и нарушения их ортогональности.

Если $\Delta f_р > \text{opt}\,\Delta f_р$, то условия различения сигналов s_1 и s_2 по сравнению с оптимальным разносом не улучшаются, а использование отведенного диапазона частот ухудшается, так как на передачу того же количества информации затрачивается более широкая полоса. При-

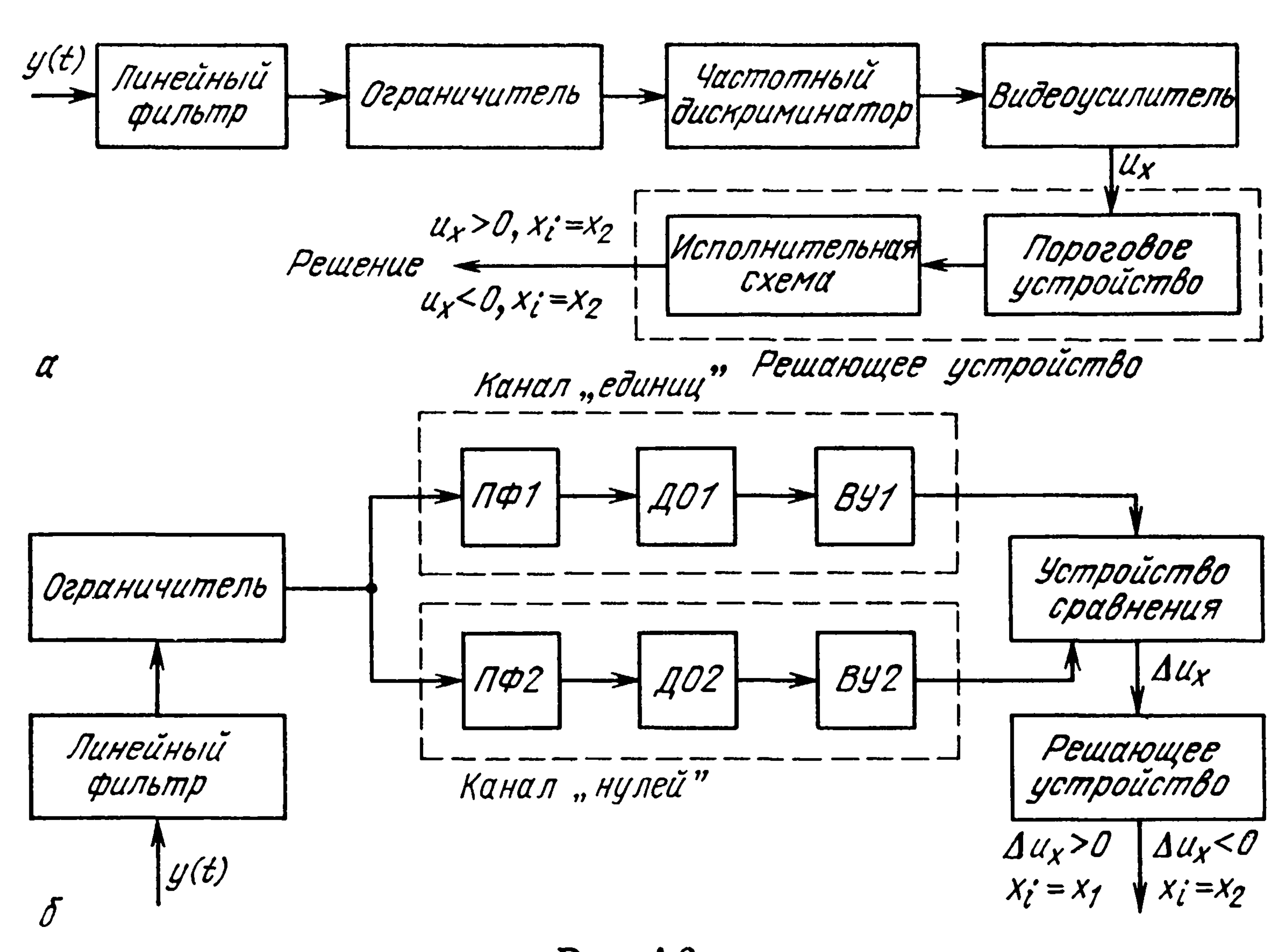


Рис. 4.6.

близительно можно считать, что минимально допустимая полоса частот, занимаемая спектром простых ЧМн сигналов, равна

$$\Delta f_э = \Delta f_р + 1/\tau_0 \approx 2/\tau_0. \qquad (4.39)$$

Некогерентный прием простых ЧМн сигналов можно осуществить несколькими способами. На рис. 4.6,*а* и *б* приведены структурные схемы двух наиболее употребительных способов реального приема ЧМн сигналов. Работа этих схем приема не требует особых пояснений.

Применение линейного приемника с двумя полосовыми фильтрами (*ПФ1* и *ПФ2*), детекторами огибающих (*ДО1* и *ДО2*) и видеоусилителями (*ВУ1*, *ВУ2*) пред-

почтительнее, чем применение нелинейного приемника с дискриминатором по следующим причинам: 1) линейный приемник проще в настройке и менее критичен к изменению параметров отдельных узлов и элементов; 2) при одинаковой полосе пропускания до ограничителя линейный приемник обеспечивает более высокую помехоустойчивость; 3) межсимвольные искажения в линейном приемнике меньше, чем в нелинейном, так как в последнем имеется дополнительная причина искажений, связанная с нелинейностью характеристики дискриминатора.

Поэтому ниже рассмотрим некогерентный прием простых ЧМн сигналов только линейным приемником с двумя фильтрами (рис. 4.6,*б*). Будем считать, что энергия посылок, полоса пропускания фильтров *ПФ1* и *ПФ2* и усиление каналов «нулей» и «единиц» одинаковы. При таких условиях схема симметрична для приема нулей и единиц и, следовательно,

$$p(s_1|s_2)=p(s_2|s_1). \qquad (4.40)$$

Для определения вероятности ошибки достаточно найти одну из этих условных вероятностей. Пусть передан сигнал s_1. Тогда ошибка произойдет, если значение огибающей $U_ш$ в канале нулей (где нет сигнала) достигнет или превысит значение огибающей $U_{сш}$ в канале единиц (где есть сигнал s_1) в момент времени, соответствующий окончанию посылки s_1.

Вероятность события $U_ш \geqslant U_{сш}$ равна

$$p(U_ш \geqslant U_{сш}) = \int\limits_{U_{сш}}^{\infty} W(U_ш)\, dU_ш. \qquad (4.41)$$

Величина $U_{сш}$ случайна и может принять с некоторой вероятностью любое значение в интервале $0 \leqslant U_{сш} < < \infty$. Поэтому для определения условной вероятности ошибки необходимо вероятность, определяемую выражением (4.41), усреднить по всем значениям случайной величины $U_{сш}$. Следовательно,

$$p(s_2|s_1) = \int\limits_0^{\infty} W(U_{сш})\, p(U_ш \geqslant U_{сш})\, dU_{сш},$$

или, с учетом выражения (4.41),

$$p(s_2|s_1)=\int\limits_0^\infty \int\limits_{U_{\text{сш}}}^\infty W(U_{\text{сш}})\,W(U_{\text{ш}})\,dU_{\text{ш}}dU_{\text{сш}}. \qquad (4.42)$$

Подставив сюда распределения огибающих, определяемые выражениями (4.7) и (4.8), и заменив переменные в соответствии с (4.9), получим

$$p(s_2|s_1)=\exp(-a^2/2)\int\limits_0^\infty v\exp(-v^2)\,I_0(av)\,dv.$$

Учитывая, что интеграл, входящий в это выражение, равен [10].

$$\int\limits_0^\infty v\exp(-v)\,I_0(av)\,dv=0{,}5\exp(a^2/4),$$

имеем

$$p(s_2|s_1)=0{,}5\exp(-a^2/4).$$

Приняв во внимание, что $a=\sqrt{2}h$, можно записать в окончательном виде формулу для вероятности ошибки при некогерентном приеме простых ЧМн сигналов

$$p_{\text{ош}}=0{,}5\exp(-h^2/2). \qquad (4.43)$$

Здесь величина h так же, как и в предыдущем случае, определяет отношение сигнал/шум в полосе, занимаемой спектром одной посылки:

$$h^2=\frac{S^2_0}{2\sigma^2}=\frac{S^2_0\tau_0}{2N_0\Delta f_{\text{ф}}\tau_0}=\frac{E}{N_0}\frac{1}{\Delta f_{\text{ф}}\tau_0}.$$

Напомним, что $\Delta f_{\text{ф}}$ — полоса пропускания фильтра нулей или единиц; E — энергия одной посылки сигнала.

При условии $\Delta f_{\text{ф}}\tau_0=1$ имеем $h^2=E/N_0=h^2_0$.

Сравним помехоустойчивость некогерентного приема ЧМн сигналов с помехоустойчивостью оптимального приема. Вероятность ошибки при оптимальном приеме ЧМн сигналов определяется выражением (3.104). Для $h_0>3$ это выражение можно представить в виде

$$p_{\text{ош}}=\frac{1}{\sqrt{2\pi}h_0}\exp\left(-\frac{h^2_0}{2}\right). \qquad (4.44)$$

Так же, как и при некогерентном приеме АМн сигналов, сравнение проведем для двух условий:

1) $h_{\text{ор}} = h_0$ — равенство энергетических затрат при передаче одного символа; при этом различие в вероятностях ошибок определяется величиной

$$\gamma_p = \frac{p_{\text{ош нк}}}{p_{\text{ош опт}}} = \sqrt{\frac{\pi}{2}}\, h_0 \approx 1{,}26 h_0. \qquad (4.45)$$

2) $p_{\text{ош нк}} = p_{\text{ош опт}}$ — равенство вероятностей ошибок; при таком условии различие в энергетических затратах характеризуется величиной

$$\gamma_h = h^2_{\text{ор}}/h^2_0 = 1 + 2\ln(1{,}26 h_0)/h^2_0. \qquad (4.46)$$

Формулы (4.45) и (4.46) характеризуют величину проигрыша, который появляется при переходе от оптимального к некогерентному приему. Расчеты показывают, что для обеспечения $p_{\text{ош}} = 10^{-3}$—10^{-6} некогерентный прием ЧМн сигналов требует увеличения энергии сигнала на 30—15% по сравнению с оптимальным, т. е. на 1—0,5 дБ.

4.4. НЕКОТОРЫЕ ОСОБЕННОСТИ ПРИЕМА ПРОСТЫХ ФАЗОМАНИПУЛИРОВАННЫХ СИГНАЛОВ

В гл. 3 было установлено, что для передачи цифровой информации оптимальны противоположные двоичные сигналы. Наиболее характерным примером таких сигналов является простой фазоманипулированный сигнал с манипуляцией фазы на π. Применение ФМн сигналов открывает принципиальные возможности для передачи цифровой информации с требуемой достоверностью при меньших затратах полосы и энергии по сравнению с другими двоичными сигналами. При ФМн информация о передаваемой цифровой последовательности заложена в фазе сигнала, который на интервале времени 0—τ_0 может принимать лишь одно из двух возможных значений:

$$s(t) = \begin{cases} s_1(t) = S_0 \sin(\omega_0 t + \varphi_1), & \text{если } x = x_1, \\ s_2(t) = S_0 \sin(\omega_0 t + \varphi_2), & \text{если } x = x_2. \end{cases} \qquad (4.47)$$

Для противофазной манипуляции

$$|\varphi_1 - \varphi_2| = \pi.$$

Выражение (4.47) можно записать и иным образом:

$$s(t) = S_0 \sin(\omega_0 t + \varphi_x), \quad 0 \leqslant t \leqslant \tau_0, \tag{4.48}$$

где φ_x — фаза, которая может принимать значения 0 или π;

$$\varphi_x = \begin{cases} \varphi_1 = 0, & \text{если } x = x_1, \\ \varphi_2 = \pi, & \text{если } x = x_2. \end{cases}$$

При передаче двоичной цифровой последовательности фаза φ_x является случайной величиной, однозначно связанной со случайной последовательностью передаваемых информационных символов x.

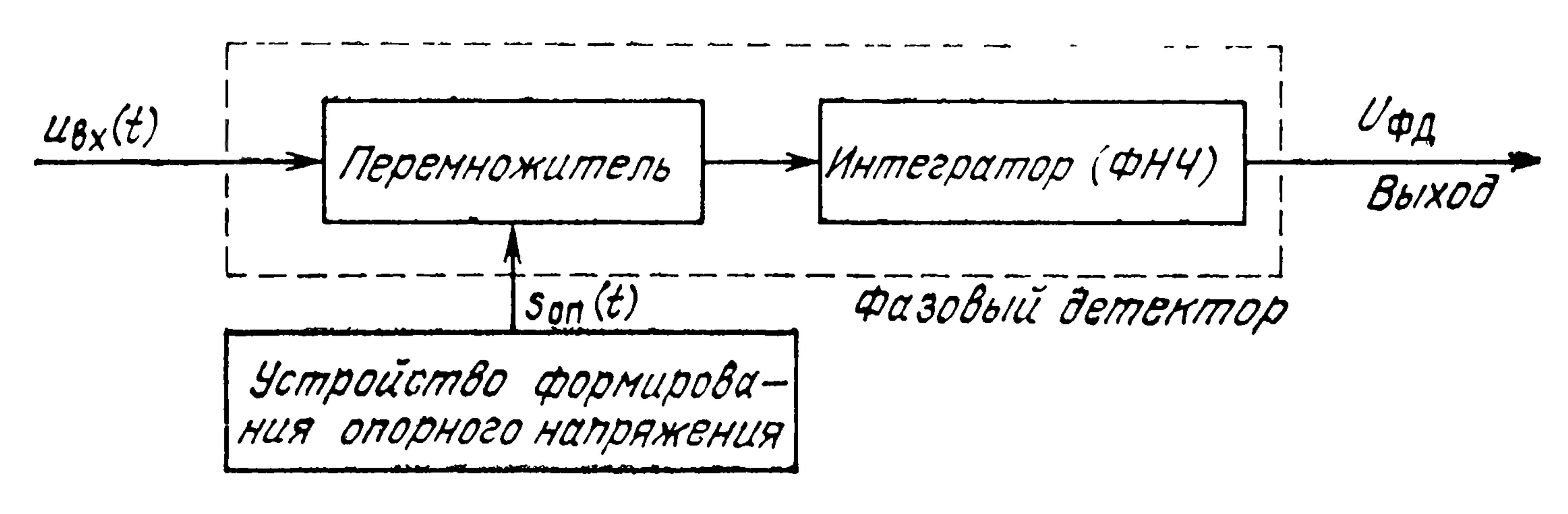


Рис. 4.7.

Демодуляция ФМн сигналов основана на возможности их фазового различения (фазовой селекции) с помощью детектора, чувствительного к фазе сигнала. Такой детектор называется фазовым. Операции, которые должен выполнять этот детектор, сходны с операциями, выполняемыми коррелятором оптимального приемника различения двух сигналов. Фазовый детектор (ФД) должен осуществлять перемножение колебания, поступившего на его вход, с опорным напряжением $s_{оп}(t)$ и фильтрацию полученного выходного напряжения от высокочастотных составляющих (рис. 4.7). Такая фильтрация выполняется благодаря частотным свойствам самой нагрузки фазового детектора, которая обычно представляет собой фильтр нижних частот (ФНЧ).

Для уяснения особенностей фазового детектора и требований к опорному напряжению рассмотрим его работу без помех. Запишем опорное напряжение в виде

$$s_{оп}(t) = S_{оп} \sin(\omega_{оп} t + \varphi_{оп}).$$

Напряжение на выходе фазового детектора с учетом фильтрации высокочастотных составляющих определяется выражением *

$$U_{ФД}=k_{ФД}\overline{s_{вх}(t)\,s_{оп}(t)}=k_{ФД}\frac{S_0S_{оп}}{2}\cos\left[(\omega_0-\omega_{оп})\,t+(\varphi_x-\varphi_{оп})\right],$$

где $k_{ФД}$ — коэффициент передачи фазового детектора, имеющий размерность [1/В]; для простоты далее будем считать $k_{ФД}=1$.

Для того чтобы напряжение на выходе фазового детектора зависело только от значения фаз сигнала и опорного напряжения, необходимо выполнять условие синхронности частоты опорного напряжения с частотой ФМн сигнала:

$$\omega_0=\omega_{оп}. \tag{4.49}$$

Условимся, что положительное напряжение на выходе детектора соответствует передаче информационного символа $x_1=1$, а отрицательное — символа $x_2=0$. Тогда с учетом условия (4.49) имеем

$$U_{ФД}=\begin{cases}+0{,}5S_0S_{оп}, & \text{если } \varphi_x-\varphi_{оп}=0,\\ -0{,}5S_0S_{оп}, & \text{если } \varphi_x-\varphi_{оп}=\pi.\end{cases} \tag{4.50}$$

Итак, опорное напряжение должно быть синхронным с сигналом и иметь начальную фазу $\varphi_{оп}=\varphi_1=0$. Фазовый детектор, опорное напряжение которого удовлетворяет такому условию, часто называют синхронным или когерентным детектором.

Нетрудно видеть, что если начальная фаза $\varphi_{оп}$ по каким-либо причинам изменит свое значение на $+\pi$ или $-\pi$, то полярность напряжений $U_{ФД}$, соответствующих посылкам сигнала s_1 и s_2, изменится на обратную. Это явление называется обратной работой фазового детектора или переходом в «негатив». Очевидно, что такой режим работы фазового детектора недопустим, поскольку с момента, соответствующего переходу в «негатив», последовательность символов будет приниматься неправильно.

Основная трудность реализации демодулятора противофазных ФМн сигналов состоит в создании синхронного опорного напряжения, имеющего неизменную на-

* Высокочастотные составляющие с частотой $\omega_0+\omega_{оп}$ фильтруются в нагрузке фазового детектора.

чальную фазу. Возможны два пути решения этой проблемы: 1) использование в приемнике местного высокостабильного генератора опорного напряжения, фазируемого с напряжением опорного генератора передатчика в начале каждого сеанса связи; 2) формирование опорного напряжения с помощью сигнала, принимаемого совместно с помехами.

В первом случае необходимо создать условия, при которых отсчет фазы передаваемых и принимаемых ФМн сигналов ведется от одного и того же значения фазы опорных генераторов, находящихся в передатчике и приемнике. Для этого в начале каждого сеанса связи нужно сфазировать указанные генераторы. Как известно, частота и фаза любого генератора не строго постоянны и под действием ряда дестабилизирующих факторов (изменения температуры окружающей среды, изменения питающих напряжений, старения деталей и т. п.) непрерывно и случайно изменяются в некоторых пределах. Поэтому через некоторый интервал времени после фазирования фазы генераторов разойдутся и прием ФМн сигналов станет невозможным.

Расчеты, выполненные в работе [11], показывают, что для реализации такого способа создания опорного напряжения требуется очень высокая стабильность частоты опорных генераторов даже для весьма коротких сеансов связи. Так, например, при относительной стабильности частоты в пределах $\delta f/f \approx 10^{-6}$—$10^{-10}$ можно обеспечить длительность сеанса связи только в пределах от $2 \cdot 10^{-3}$ до 20 с. Обеспечить требуемую стабильность даже для коротких сеансов связи трудно. Кроме того, применение независимых генераторов опорных напряжений на передаче и приеме исключает учет сравнительно небольших случайных изменений частоты и фазы сигнала, которые неизбежны в реальных каналах связи при распространении сигналов.

Отмеченные особенности делают невозможным использование независимых генераторов в существующих линиях связи. Однако говорить о полной бесперспективности рассмотренного метода формирования опорного напряжения нельзя. В настоящее время значительное внимание уделяется исследованиям вопросов передачи информации по волноводным линиям связи. Такие линии связи обладают рядом значительных достоинств: огромной пропускной способностью, высокой

стабильностью характеристик, малым уровнем помех. Волноводные линии связи не создают помех для других систем и не подвержены помехам, создаваемым работой других систем, т. е. такие линии связи полностью решают проблему электромагнитной совместимости различных систем, использующих одни и те же диапазоны частот. Высокое постоянство характеристик волноводных линий связи и малый уровень помех в них открывают определенные возможности для применения в этих линиях квантовых опорных генераторов, обладающих очень высокой стабильностью частоты (от 10^{-11} до 10^{-13}).

В существующих линиях связи опорное напряжение формируется с помощью принимаемого сигнала. Сформированное напряжение не должно иметь изменений фазы и быть по возможности свободным от шумов и других помех. В настоящее время известен ряд способов и соответствующих им устройств формирования опорного напряжения (УФОН). Все эти способы основаны на двух операциях: 1) снятие манипуляции с ФМн сигнала и 2) фильтрация полученного напряжения от помех.

4.5. СХЕМЫ ФОРМИРОВАНИЯ ОПОРНОГО НАПРЯЖЕНИЯ В ПРИЕМНИКАХ ФМн СИГНАЛОВ

В литературе [11, 12] описано значительное число различных схем формирования опорного напряжения, предложенных для реализации фазочувствительных детекторных устройств в приемниках ФМн сигналов. Ниже рассматриваются некоторые схемы, получившие большую известность и достаточно широкое применение.

4.5.1. Схема Пистолькорса

Эта схема, приведенная на рис. 4.8, была предложена советским ученым А. А. Пистолькорсом в 1933 г. для фазовой телеграфии [11].

Нетрудно убедиться в том, что удвоение частоты сигнала, манипулированного по фазе на π, позволяет получить на выходе удвоителя (×2) колебание с неизменной начальной фазой. Пропустив это колебание через узкополосный резонансный фильтр (*УРФ*) с частотой настройки $2f_0$, можно получить колебание, близкое к синусоидальному. При достаточно узкой полосе фильтра уровень помех на его выходе значительно меньше уровня помех на входе основного фазового детектора (*ФД*). После деления частоты на два (:2) получаем необходимое опорное напряжение с частотой f_0 без манипуляции фазы и с малым уровнем помех. Так как в тракте форми-

рования опорного напряжения появляются дополнительные фазовые сдвиги (особенно в узкополосном фильтре), для их компенсации применяется фазовращатель (*ФВ*), сдвигающий в противоположном направлении фазу опорного напряжения на величину появляющихся фазовых сдвигов.

Заметное ослабление уровня помех в тракте формирования опорного напряжения по сравнению с уровнем помех в основном тракте достигается при соотношении $\Delta f_ф/\Delta f_э \approx \Delta f_ф \tau_0 \leqslant 0{,}1$. Однако для выполнения такого условия необходимо значительно повышать требования к стабильности частоты сигнала, а также налагать определенные ограничения на скорость случайных изменений фазы сигнала в среде распространения. Поэтому при заметной нестабильности частоты сигнала или значительных ее изменениях за счет эффекта Допплера, вызванных перемещениями передатчика или приемника (или обоих вместе) в пространстве, пассивный узкополосный фильтр заменяют активным следящим фильтром, отслеживающим изменения частоты сигнала. Существует ряд схем таких фильтров. Все они основаны на принципе частотной (ЧАП) или фазовой (ФАП) автоподстройки частоты [11, 12].

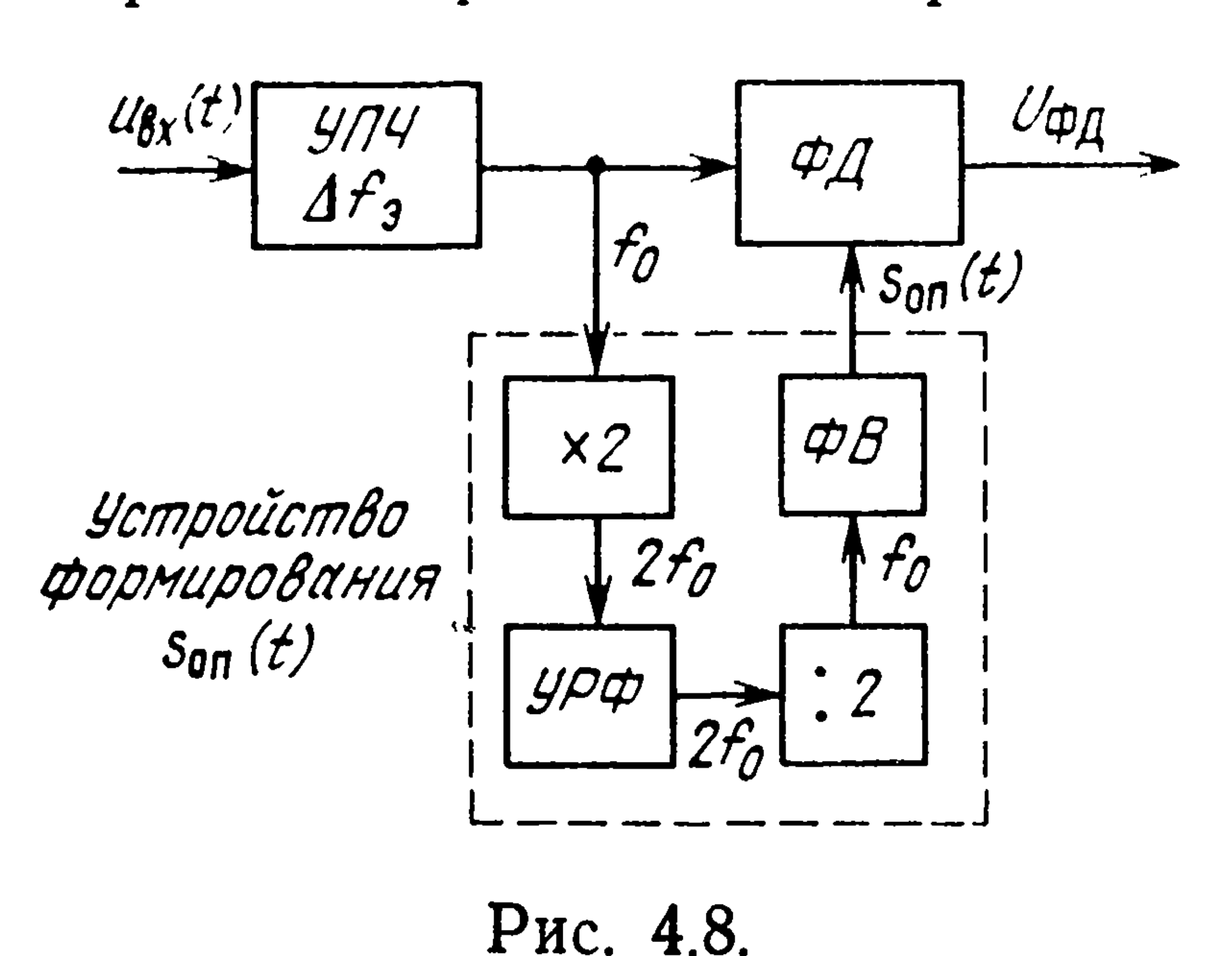


Рис. 4.8.

Исследование показывает [11], что фаза опорного напряжения, сформированного в схеме Пистолькорса, не однозначна и в зависимости от начальных условий может принимать одно из двух возможных устойчивых состояний, отличающихся друг от друга на 180°. Как указывалось выше, такое положение приводит к появлению «обратной работы» основного фазового детектора, в результате которой полярности информационных символов на выходе фазового детектора изменяются на обратные. Помимо указанного обстоятельства, причиной «обратной работы» могут быть различные возмущения (помехи в тракте опорного напряжения, переходные процессы в схеме), действие которых приводит к перескоку фазы опорного напряжения из одного устойчивого состояния в другое. Наличие делителя частоты в схеме Пистолькорса увеличивает неопределенность начальной фазы опорного напряжения и возможность перескока из одного состояния в другое.

4.5.2. Схема Сифорова

Схема, приведенная на рис. 4.9, предложена советским ученым В. И. Сифоровым в 1937 г. В этой схеме манипуляция с сигнала снимается удвоением частоты сигнала в удвоителе (*УДВ1*), а фильтрация помех в канале формирования опорного напряжения обеспечивается фазовой автоподстройкой частоты (*ФАП*) управляемого генератора (*УГ*). Схема не содержит делителя частоты. Это достигается тем, что сравнение в фазовом дискриминаторе (*ДФ*) выполняется на удвоенной частоте. Однако отсутствие делителя не устраняет неоднозначности фазы полученного опорного напряжения.

Если эта фаза по каким-либо причинам изменится на 180°, то схема *ФАП* не отреагирует на это, поскольку фаза на выходе *УДВ2* остается без изменения. В результате основной фазовый детектор перейдет в режим «обратной работы».

Фазовращатель в схеме рис. 4.9 служит для компенсации фазовых сдвигов, вносимых различными элементами схемы в фазу напряжения управляемого генератора.

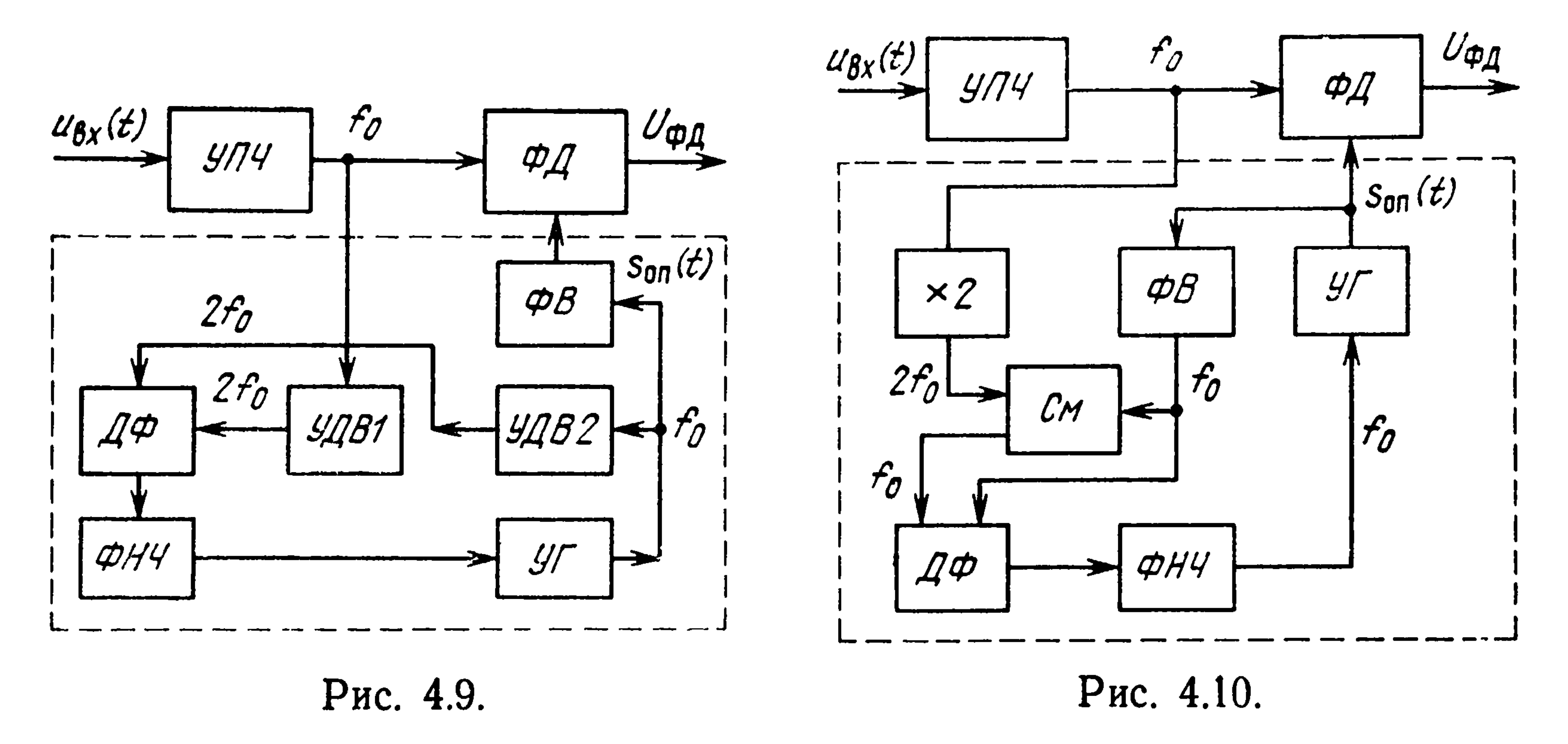


Рис. 4.9. Рис. 4.10.

Схема Сифорова удобнее в реализации, чем схема Пистолькорса, и менее подвержена перескокам фазы опорного напряжения.

Наличие удвоителя частоты управляемого генератора в рассмотренной схеме Сифорова не является принципиально необходимым. Возможен вариант схемы, в которой удвоитель частоты в системе *ФАП* отсутствует. Такая схема приведена на рис. 4.10. Как и в предыдущих схемах, манипуляция с ФМн сигнала снимается удвоением частоты сигнала. Однако фазовый дискриминатор системы *ФАП* в этой схеме работает на промежуточной частоте сигнала. Для этого применяется дополнительный смеситель (*См*), роль напряжения гетеродина в котором выполняет напряжение управляемого генератора (*УГ*). Исключение удвоителя частоты из системы *ФАП* благоприятно сказывается на свойствах схемы и уменьшает вероятность перескоков фазы опорного напряжения.

4.5.3. Схема Костаса

В 1956 г. американский ученый Д. Костас предложил схему формирования опорного напряжения для синхронного (когерентного) детектирования амплитудно-модулированных сигналов [11, 12].

Эта схема представлена на рис. 4.11. Влияние манипуляции фазы сигнала устраняется в ней перемножителем, на один вход которого поступает сигнал с выхода усилителя-ограничителя, а на второй — сигнал с выхода фазового детектора (*ДФ*) системы *ФАП*. По сравнению с ранее рассмотренными схемами формирования такой способ снятия манипуляции обеспечивает более высокую помехоустойчивость схемы к флюктуационным помехам. Недостатком схемы является некоторая ее сложность, обусловленная тем, что перемножитель должен работать на постоянном токе. Это приво-

дит к необходимости пропускания постоянных и очень медленно меняющихся напряжений, вплоть до управителя частоты. При этом входы перемножителя оказываются связанными гальванически и для их взаимной развязки необходимо применять мостовые схемы постоянного тока.

Этот недостаток может быть устранен, если снятие манипуляции фазы сигнала осуществлять перемножителем, работающим на частоте сигнала. Вариант подобной схемы показан на рис. 4.12. Схема работает следующим образом. Последовательность посылок ФМн сигнала в перемножителе умножается на разнополярную по-

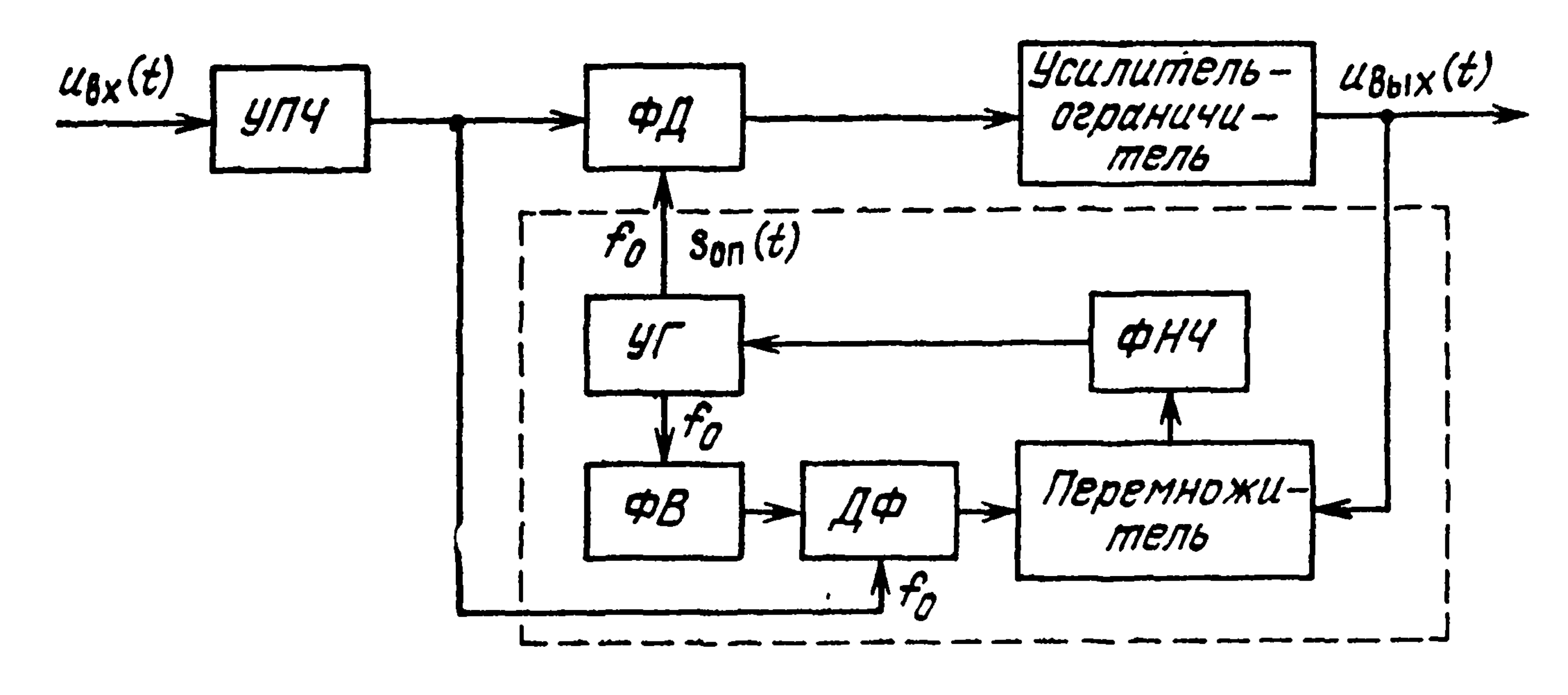


Рис. 4.11.

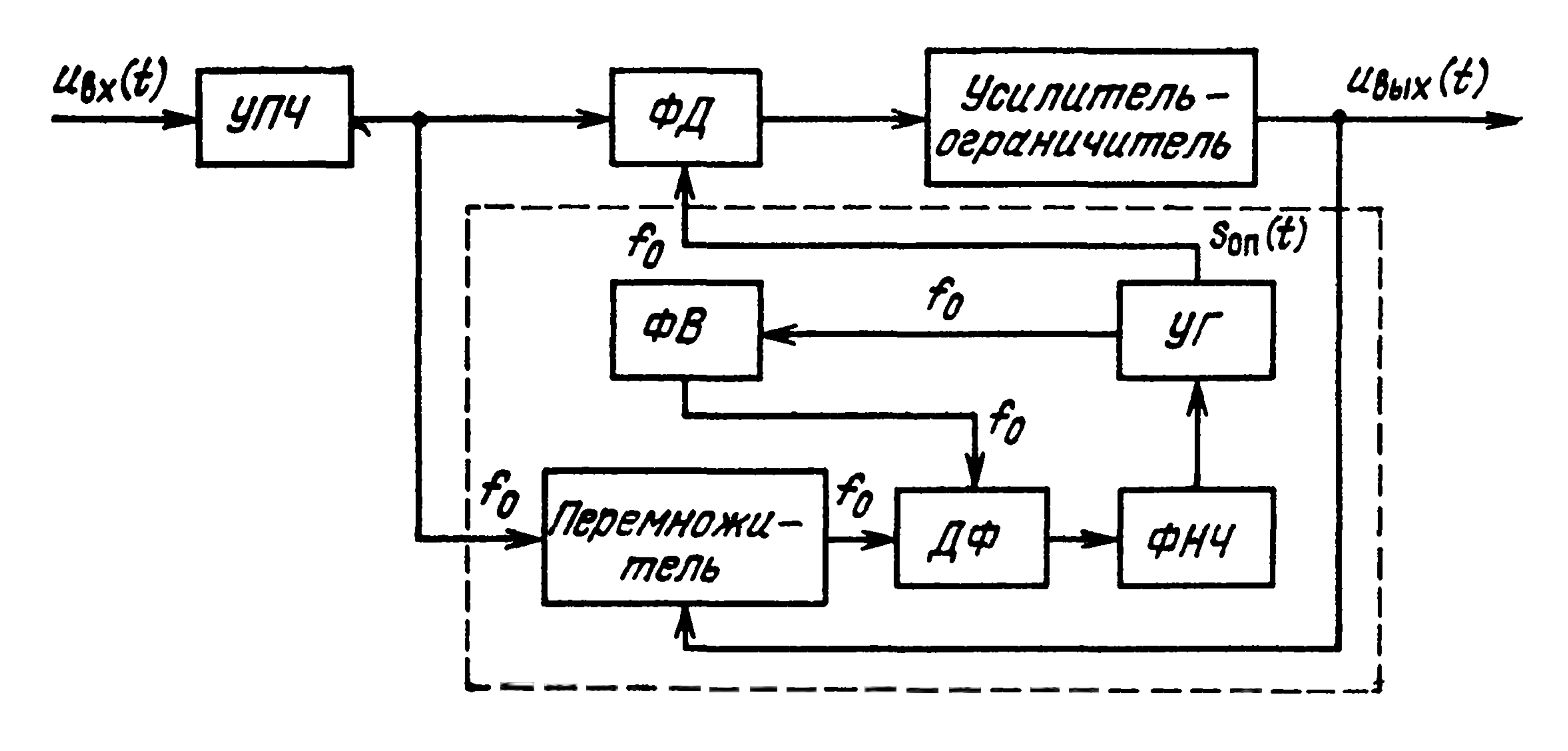


Рис. 4.12.

следовательность импульсов, соответствующих информационным символам на выходе усилителя-ограничителя. В результате этого достигается снятие манипуляции с сигнала. Полученное напряжение, имеющее частоту сигнала, поступает на вход фазового детектора системы *ФАП*, с помощью которой подстраиваются частота и фаза управляемого генератора, вырабатывающего опорное напряжение. Этот вариант схемы проще в реализации, но его помехоустойчивость ниже, так как влияние помех в тракте опорного напряжения проявляется сильнее, чем в первой схеме.

Так же, как и предыдущие схемы, схема Костаса имеет склонность к переходу в режим «обратной работы», так как в ней также

возможны два устойчивых значения начальной фазы опорного напряжения, отличающиеся на 180°.

Помимо рассмотренных схем, известен ряд других схем формирования опорного напряжения для детектирования противофазных ФМн сигналов [11, 12, 14]. Всем им присущ уже отмеченный выше принципиальный недостаток, проявляющийся в виде эффекта «обратной работы» фазового детектора. Этот недостаток был основным препятствием на пути практической реализации каналов связи с ФМн сигналами.

В обычных неавтоматизированных линиях телеграфной связи, где получателем был человек-оператор, с эффектом «обратной работы» иногда можно было мириться, проводя перед каждым сеансом связи предварительную регулировку аппаратуры, а также устанавливая правильную фазировку опорного напряжения каждый раз, когда возникал эффект «обратной работы» или появлялся перерыв в связи.

В современных автоматизированных линиях передачи дискретной информации, где требуются высокие скорости передачи, а получателем информации является автоматизированное устройство или ЭЦВМ, режим «обратной работы» недопустим.

Таким образом, практическое осуществление приема классических оптимальных (т. е. с манипуляцией фазы на 180°) ФМн сигналов с использованием фазового детектора наталкивается на принципиально неустранимую трудность, обусловленную «обратной работой» этого детектора. Причина заключается в том, что при манипуляции фазы сигнала на 180° сам ФМн сигнал содержит в себе неопределенность начальной фазы, так как в спектре такого сигнала отсутствует составляющая на частоте несущей. Поэтому, даже имея в приемном устройстве идеальное опорное напряжение, точно совпадающее по частоте с сигналом, можно принять либо истинную последовательность переданных символов, либо обратную (инверсную) ей в зависимости от начальных условий фазировки входного сигнала и опорного напряжения.

Положение усугубляется тем, что даже при правильной начальной фазировке в реальных каналах связи всегда имеются причины, вызывающие спонтанный перескок фазы опорного напряжения, приводящий к «обратной работе» фазового детектора. К числу таких причин относятся помехи в канале связи, переходные процессы при прохождении ФМн сигналов через тракт приемника, изменение питающих напряжений и ряд других воздействий, учесть которые практически невозможно.

4.6. МЕТОД ПЕРЕДАЧИ ЦИФРОВОЙ ИНФОРМАЦИИ СИГНАЛАМИ С ОТНОСИТЕЛЬНОЙ ФАЗОВОЙ МАНИПУЛЯЦИЕЙ

В связи с невозможностью устранения явления обратной работы фазового детектора были предложены способы, которые основывались на отказе от классического вида противофазных ФМн сигналов: введение в пе-

редаваемые ФМн сигналы специальных пилот-сигналов, с помощью которых формируется опорное напряжение; применение отдельных посылок для фазирования генератора опорного напряжения в приемнике; уменьшение угла манипуляции для сохранения в спектре ФМн сигнала составляющей на частоте сигнала, которая выполняет роль пилот-сигнала и содержит информацию о начальной фазе несущей. Несмотря на то, что перечисленные способы позволяют устранить эффект «обратной работы», они находят применение только для решения отдельных частных задач, так как заметно снижают предельные возможности передачи информации сигналами с фазовой манипуляцией и требуют усложнения аппаратуры систем.

В литературе известны также предложения, которые сводятся к отказу от фазового детектирования ФМн сигналов и применению различных частотных детекторов для демодуляции ФМн сигналов [11]. В таких детекторах «обратная работа» отсутствует, однако и помехоустойчивость предлагаемых способов приема значительно ниже, что в ряде случаев сводит на нет целесообразность применения ФМн сигналов.

Итак, стремление как можно полнее реализовать преимущества противофазных ФМн сигналов наталкивается на принципиальную трудность, связанную с эффектом «обратной работы» фазового детектора, а стремление обойти эту трудность приводит к ухудшению качественных показателей канала связи по сравнению с предельно возможными для ФМн сигналов.

В 1954 г. советский ученый Н. Т. Петрович предложил метод передачи, который свободен от эффекта «обратной работы» и в то же время позволяет достаточно полно реализовать преимущества фазовой манипуляции в реальных каналах связи. Метод был предназначен для фазовой телеграфии и назван «относительной фазовой телеграфией» (ОФТ) или «относительной фазовой манипуляцией» (ОФМн). Последний термин получил наибольшее распространение в отечественной литературе [11, 12]*.

---

* Помимо наиболее распространенных терминов ОФТ и ОФМн в литературе этот метод встречается также под названием ФРМн (фазоразностная манипуляция) и «дифференциальная когерентная система» (в иностранной литературе).

Рассмотрим суть этого метода. Пусть имеется исходная двоичная последовательность информационных символов, которую необходимо передать по каналу связи. Для передачи этой последовательности необходимо иметь однозначно связанную с ней последовательность посылок сигналов. При ФМн сигнале посылки имеют вид, определяемый выражением (4.47).

При классической противофазной ФМн между информационными символами и посылками сигнала должно быть следующее соответствие:

$$x = x_1 \rightarrow s(t) = s_1(t),\ \varphi_1 = 0;$$
$$x = x_2 \rightarrow s(t) = s_2(t),\ \varphi_2 = \pi. \qquad (4.51)$$

В этом случае определение фазы любой принятой посылки производится по отношению к некоторой фиксированной опорной фазе, за которую принимается в данном случае $\varphi_{оп} = \varphi_1 = 0$.

Н. Т. Петрович предложил отказаться от абсолютной системы отсчета фазы и ввести относительную (скользящую) систему отсчета фазы для каждой принимаемой посылки. Началом отсчета фазы каждой очередной посылки является фаза предыдущей посылки.

Пусть φ_j — фаза j-й посылки, а φ_{j-1} — фаза $(j-1)$-й посылки. Так как противофазный ФМн сигнал может принимать лишь два значения с фазами $\varphi_1 = 0$ и $\varphi_2 = \pi$, то разность фаз между двумя соседними посылками может также принимать только два значения

$$\Delta\varphi = |\varphi_j - \varphi_{j-1}| = \begin{cases} 0, & \text{если } \varphi_j = \varphi_{j-1}, \\ \pi, & \text{если } \varphi_j = \varphi_{j-1} + \pi. \end{cases}$$

Введение относительной системы отсчета фаз посылок приводит к тому, что соответствие между информационными символами и посылками сигналов становится иным по сравнению с классической ФМн. При относительном методе передачи противофазных ФМн сигналов выбор сигнальной посылки зависит как от вида информационного символа ($x_1 = 1$ или $x_2 = 0$), поступающего на вход модулятора, так и от того, какой была предыдущая посылка (s_1 или s_2). Если условиться, что символу $x_1 = 1$ должна соответствовать посылка сигнала, для которой $\Delta\varphi = 0$, а символу $x_2 = 0$ посылка, для которой $\Delta\varphi = \pi$, то получим следующее правило манипуляции фазы несущего колебания при передаче двоичной инфор-

мации методом ОФМн: при передаче символа 1 фаза посылки остается такой же, как и у предыдущей посылки, а при передаче символа 0 фаза посылки изменяется на 180° по отношению к фазе предыдущей посылки. Условно это правило можно записать так:

$$x_j = \begin{cases} x_1 = 1 \rightarrow \varphi_j = \varphi_{j-1}, \\ x_2 = 0 \rightarrow \varphi_j = \varphi_{j-1} + \pi, \end{cases} \qquad (4.52)$$

где j — номер посылки, или выразить в виде табл. 4.1.

Таким образом, при передаче методом ОФМн одно и то же значение информационного символа может быть передано как сигналом $s_1(t)$, так и сигналом $s_2(t)$, в зависимости от того, каков был предыдущий сигнал *. Очевидно, что в передающем устройстве системы с ОФМн необходимо перед началом каждого сеанса связи обеспечивать создание вспомогательной посылки, от фазы которой начинается отсчет фазы первой информационной посылки сигнала. Передающее устройство должно иметь элемент памяти для запоминания фазы каждой посылки, чтобы обеспечить начало отсчета фазы следующей посылки.

Таблица 4.1

Символ	Посылка	
	$s_{j-1}(t)$	$s_j(t)$
$x = 1$	$s_1(t)$ $s_2(t)$	$s_1(t)$ $s_2(t)$
$x = 0$	$s_1(t)$ $s_2(t)$	$s_2(t)$ $s_1(t)$

В приемном устройстве ОФМн сигналов нельзя принимать решение только по одной принятой посылке. Так как при ОФМн передаваемая информация заложена не в самом значении фазы данной посылки сигнала, а в разности фаз данной и предыдущей посылок, то для решения о том, какой информационный символ был передан, необходимо учитывать не только данную посылку, но и предыдущую принятую. Следовательно, при передаче информации методом ОФМн необходимо располагать в передатчике и приемнике элементами памяти

* Рассмотренное правило манипуляции фазы при ОФМн может быть и иным: при передаче символа 1 фаза посылки меняется на обратную по отношению к фазе предыдущей посылки, а при передаче символа 0 остается без изменения. Выбор того или иного правила определяется удобствами практической реализации системы с ОФМн.

для запоминания передаваемой и принятой посылок, с помощью которых обеспечивается относительное начало отсчета фаз при передаче и приеме.

Для иллюстрации рассмотренного метода в табл. 4.2 приведен пример соответствия между информационными символами и посылками сигнала при классической противоположной ФМн и ОФМн.

Таблица 4.2

Номер столбца	0	1	2	3	4	5	6	7	8	9	10	11	12
x_j	1	0	1	0	0	1	1	0	0	0	1	1	1
s_j ФМн	s_1	s_2	s_1	s_2	s_2	s_1	s_1	s_2	s_2	s_2	s_1	s_1	s_1
s_j ОФМн	s_1	s_2	s_2	s_1	s_2	s_2	s_2	s_1	s_2	s_1	s_1	s_1	s_1
x_{kj}	1	1	0	0	0	1	0	0	0	0	1	0	1
s_{kj}ОФМн	s_1	s_1	s_2	s_2	s_2	s_1	s_2	s_2	s_2	s_2	s_1	s_2	s_1

В этой таблице в выделенном столбце с номером $j=0$ указаны вспомогательный символ и соответствующая ему вспомогательная посылка, передаваемые в начале сеанса связи.

Для получения ОФМн в передающем устройстве необходимо иметь дополнительную схему, которая бы определяла границы посылок и управляла поворотом фазы этих посылок на 180° в соответствии с принятым правилом. Из рассмотрения табл. 4.2 видно, что поворот фазы посылки необходимо осуществлять как при переходе от символа 1 к символу 0, так и при переходе от символа 0 к символу 0. Практически во многих случаях более удобен режим работы, при котором передатчик работает так же, как и при классической ФМн, т. е. переброс фазы посылки происходит только при смене одного символа на другой (1 на 0 или 0 на 1). Для этого достаточно перекодировать исходную цифровую последовательность двоичных символов x_j по правилу

$$x_{\text{к}j} = x_j \oplus x_{\text{к}(j-1)} \tag{4.53}$$

и полученную новую последовательность подать на модулятор передатчика.

Соответствующая структурная схема передатчика показана на рис. 4.13. Устройство перекодирования состоит из двоичного сумматора и элемента памяти, выполняющего запоминание перекодированного символа на время, равное длительности этого символа ($\tau_з=\tau_0$).

В табл. 4.2 в двух последних строках приведены значения перекодированной последовательности символов $x_{кj}$ и соответствующие им значения посылок сигнала $s_{кj}$.

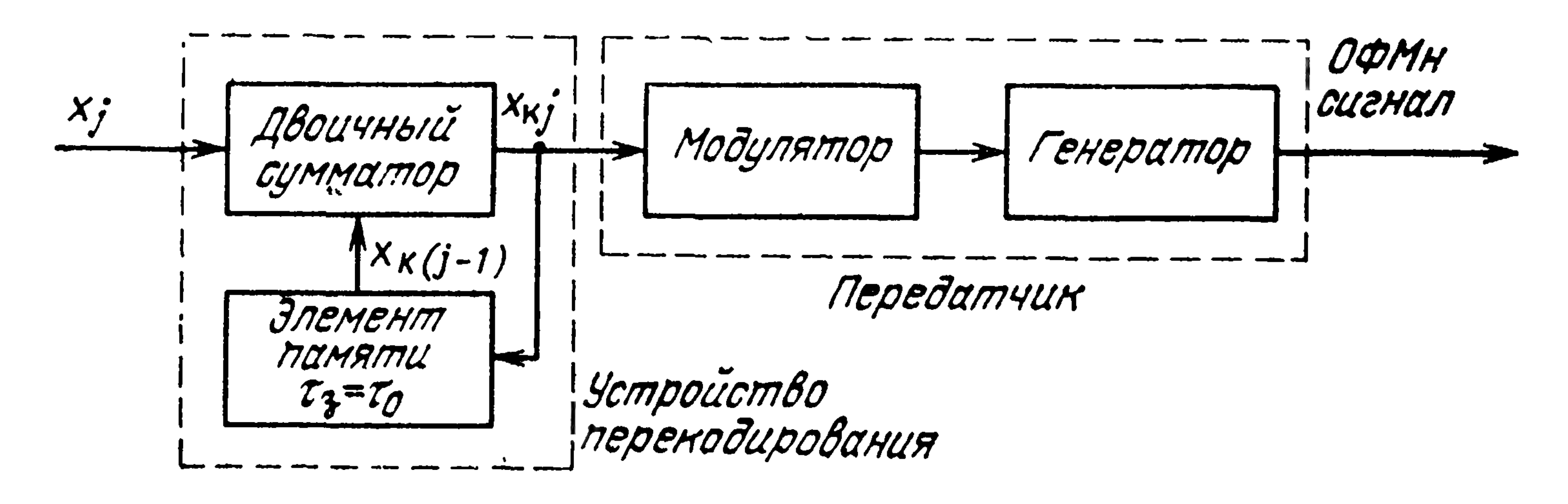


Рис. 4.13.

Естественно, что после приема переданной последовательности необходимо выполнить обратное перекодирование чтобы вернуться к исходной цифровой последовательности. Правило, по которому выполняется такое перекодирование, зависит от применяемого способа приема сигналов с ОФМн. Поэтому этот вопрос будет рассмотрен более подробно в следующем параграфе.

В настоящее время известен ряд способов формирования сигналов с ОФМн. С некоторыми из них более подробно можно ознакомиться в работах [12, 14, 15].

4.7. МЕТОДЫ ПРИЕМА СИГНАЛОВ С ОТНОСИТЕЛЬНОЙ ФАЗОВОЙ МАНИПУЛЯЦИЕЙ

В настоящее время известны и находят широкое применение два основных метода приема ОФМн сигналов: 1) корреляционный (когерентный), основанный на использовании специально сформированного когерентного опорного напряжения; 2) автокорреляционный (некогерентный), основанный на использовании принимаемых посылок сигнала в качестве опорного напряжения.

4.7.1. *Корреляционный (когерентный) метод приема*

При таком методе приема демодуляция ОФМн сигналов осуществляется фазовым детектором, опорное напряжение для которого формируется одной из рассмотренных ранее схем. Так как при формировании опорного напряжения из принимаемого сигнала действие шумовой помехи не может быть устранено полностью, то

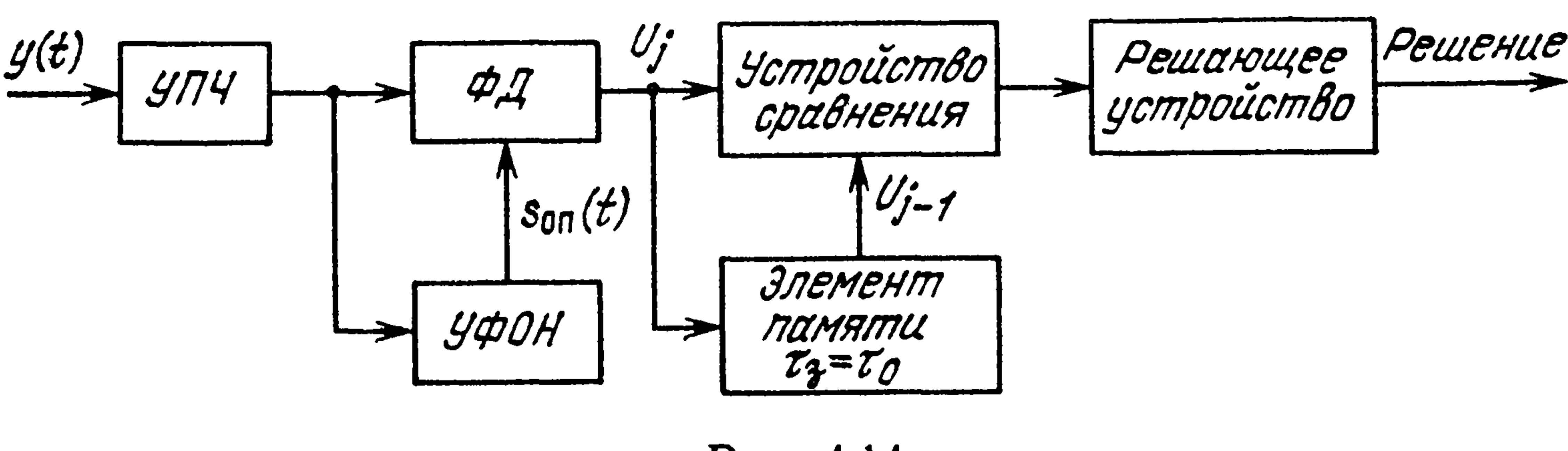


Рис. 4.14.

полученное опорное напряжение, строго говоря, не полностью когерентно с сигналом. Поэтому рассматриваемый метод приема лишь в первом приближении можно считать когерентным. По указанной причине в литературе такой метод иногда называют частично-когерентным [16]. Структурная схема, соответствующая рассматриваемому методу приема при ОФМн без перекодирования, приведена на рис. 4.14.

Решение о том, какой символ был передан, принимается при сравнении полярности огибающей принятой посылки с полярностью огибающей предыдущей посылки. Для сравнения полярностей в приемном устройстве необходимо иметь элемент памяти, запоминающий полярность предыдущей посылки. Такой элемент может быть выполнен на основе одной из спусковых схем, имеющей два устойчивых состояния (триггер, кипп-реле, фантастрон).

Устройство сравнения обычно представляет собой схему совпадения. При совпадении полярностей посылок в схеме вырабатывается управляющий импульс, под действием которого решающее устройство фиксирует наличие символа $x_1(1)$, в противном случае фиксируется наличие символа $x_2(0)$. Изложенный способ приема получил в литературе название «сравнение полярностей» [11]. Особенности когерентного приема сигналов

Таблица 4.3

ОФМн	без перекодирования												
Номер столбца j	0	1	2	3	4	5	6	7	8	9	10	11	12
x_j	1	0	1	0	0	1	1	0	0	0	1	1	1
s_j ОФМн	s_1	s_2	s_2	s_1	s_2	s_2	s_2	s_1	s_2	s_1	$\underline{s_2}$	s_1	s_1
$s_{оп}$	s_1	s_1	s_1	s_1	s_1	s_1	$\underline{s_2}$	s_2	s_2	s_2	s_2	s_2	s_2
Полярность U_j	+	−	−	+	−	−	±	−	+	−	±	−	−
Полярность $U_{(j-1)}$		+	−	−	+	−	−	+	−	+	−	±	−
$\hat{x}_j$	1	0	1	0	0	1	0	0	0	0	0	0	1
ОФМн	**с перекодированием**												
$x_{кj}$	1	1	0	0	0	1	0	0	0	0	1	0	1
$s_{кj}$ ОФМн		s_1	s_2	s_2	s_2	s_1	s_2	s_2	s_2	s_2	s_1	s_2	s_1
$s_{оп}$	s_1	s_1	s_1	s_1	s_1	s_1	s_2	s_2	s_2	s_2	s_2	s_2	s_2
Полярность U_j	+	+	−	−	−	+	±	+	+	+	−	+	−
$\hat{x}_{кj}$	1	1	0	0	0	1	1	1	1	1	0	1	0
$\hat{x}_j$	1	0	1	0	0	1	0	0	0	0	1	1	1

с ОФМн способом сравнения полярностей проиллюстрируем с помощью табл. 4.3.

В этой таблице обозначено: j — номер символа в передаваемой последовательности; x_j — вид символа (1 или 0); s_j — посылки сигнала; U_j — напряжение на выходе фазового детектора; $\hat{x}_j$ — принятый символ; $s_{оп}$ — опорный сигнал. В столбце с номером $j=0$ приведены вспомогательные символы и посылки, предшествующие началу передачи информационной последовательности символов.

Первая часть таблицы относится к приему ОФМн сигналов без перекодирования информации в передатчике. Посылки сигналов $s_j(t)$ образованы в соответствии с правилом получения ОФМн сигналов, рассмотренным в § 4.6. Прием сигналов и формирование принятой последовательности символов $\hat{x}_j$ выполняются в соответствии со структурной схемой рис. 4.14.

Нетрудно видеть, что перескок фазы опорного напряжения при приеме ОФМн сигналов способом сравнения полярностей приводит к появлению только локальной ошибки в принимаемой последовательности символов. Если скачок фазы произойдет точно на границе соседних посылок s_{j-1} и s_j, то неправильно будет принят только один символ x_j. Для примера в табл. 4.3 показано, что в момент времени, соответствующий границе между 5- и 6-й посылками, произошел скачок фазы опорного напряжения и его фаза стала соответствовать фазе посылок s_2 (вместо фазы посылки s_1).

При классической ФМн с этого же момента фазовый детектор перешел бы в негативный режим работы. В рассматриваемом же случае произойдет лишь одиночная ошибка и вместо символа $\hat{x}_j$ $(j=6)$, соответствующего единице, появится символ нуль*.

Если скачок фазы опорного напряжения произойдет в момент времени, находящийся в пределах длительности одной посылки сигнала, то импульс на выходе фазового детектора «раздробится» на две части с разной полярностью. Можно показать, что в результате такого «дробления» импульса возможен ошибочный прием двух соседних символов (особенно, если скачок фазы происходит в момент времени, близкий к середине посылки).

Таким образом, перескок фазы опорного напряжения при приеме ОФМн сигналов способом сравнения полярностей приводит к появлению ошибок локального характера, которые охватывают один или в крайнем случае два символа. Это принципиально отличает частично-когерентный прием ОФМн сигналов от приема ФМн сигналов, где скачок фазы опорного напряжения вызывает эффект обратной работы фазового детектора, который приводит к ошибочному приему всех символов до следующего скачка фазы.

---

* В табл. 4.3 для лучшего различия перескок фазы опорного напряжения и ошибочно принятый символ подчеркнуты снизу. То же относится и к изменению полярности напряжения U_j.

При приеме ОФМн сигналов, помимо рассмотренных ошибок, неизбежны также ошибки, вызванные непосредственным действием шумов на передаваемые посылки сигнала. Если под действием помехи фаза посылки изменится настолько, что это приведет к изменению полярности напряжения на выходе фазового детектора, то ошибочно будет зарегистрирован не только переданный символ, но и символ, следующий за ним.

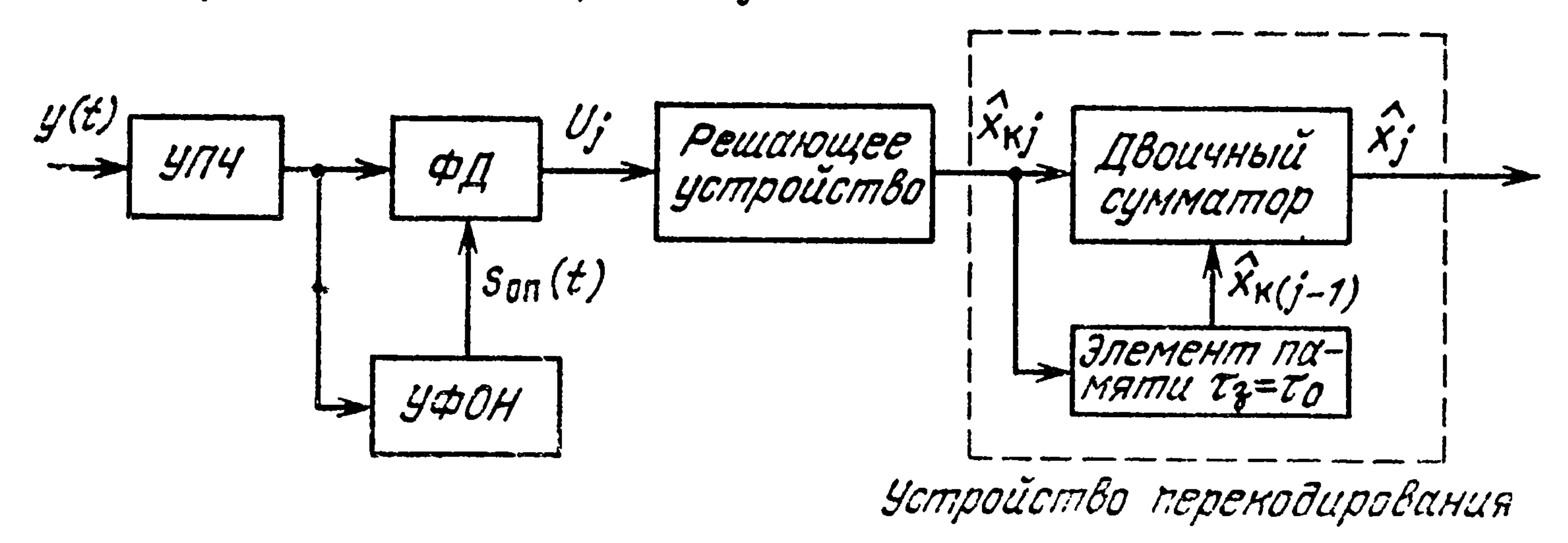


Рис. 4.15.

В табл. 4.3 сказанное иллюстрируется столбцом $j=10$. Если под действием помехи фаза посылки s_1 изменится так, что будет воспринята как фаза посылки s_2 (это условно обозначено в виде $\underline{s_2}$), то это приведет к изменению знака полярности U_j и ошибочному приему символов с номерами $j=10$ и $j=11$.

Вторая часть табл. 4.3 относится к приему ОФМн сигналов с перекодированием. Перекодирование исходной последовательности выполняется по правилу (4.53) и полученная последовательность $x_{кj}$ передается посылками $s_{кj}$, формируемыми как и при классической ФМн. Здесь символу $x_{кj}=1$ ставится в соответствие посылка $s_{кj}=s_1$, а $x_{кj}=0$ — посылка $s_{кj}=s_2$.

Прием сигналов и формирование принятой последовательности символов $\hat{x}_j$ выполняется в соответствии со структурной схемой рис. 4.15. В зависимости от полярности напряжения на выходе фазового детектора решающее устройство формирует тот или иной символ (в нашем случае положительной полярности ставится в соответствие символ 1, а отрицательной — символ 0).

Последовательность символов $\hat{x}_{кj}$ подвергается обратному перекодированию по правилу

$$\hat{x}_j = \hat{x}_{кj} \oplus \hat{x}_{к(j-1)}. \qquad (4.54)$$

Здесь $\hat{x}_{\text{к}j}$ и $\hat{x}_{\text{к}(j-1)}$ — принятые соседние символы перекодированной последовательности.

В результате на выходе схемы образуется последовательность информационных символов $\hat{x}_j$, соответствующая исходной переданной последовательности x_j. Устройство обратного перекодирования состоит из тех же элементов, что и устройство перекодирования в передатчике.

Рассмотренный способ приема отличается от классического приема противоположных ФМн сигналов только введением дополнительного устройства перекодирования. Из рассмотрения табл. 4.3 следует, что при перескоке фазы опорного напряжения (9-я строка 6-й столбец) последовательность символов $\hat{x}_{\text{к}j}$ переходит в «негатив», так же, как и при классической ФМн. Однако последующее перекодирование приводит к устранению негатива для всех символов $\hat{x}_j$, кроме символа $\hat{x}_j$ с номером $j=6$, т. е. влияние перескоков фазы опорного напряжения в этом способе приема, как и в способе сравнения полярностей, проявляется только локально. Можно показать, что искажения отдельных посылок сигнала шумами будут сказываться в рассмотренном способе так же, как и в способе сравнения полярностей, т. е. приводить к парным ошибкам в принятой последовательности символов $\hat{x}_j$. Следовательно, оба способа приема эквивалентны с точки зрения перескоков фазы опорного напряжения и действия шумов на посылки сигнала.

Таким образом, одиночные искажения посылок ОФМн сигнала шумами могут приводить к парной ошибке при приеме. Вероятность этой ошибки можно найти из следующих соображений. Ошибочная регистрация символа $\hat{x}_j$ при приеме способом сравнения полярностей может произойти в результате одного из двух несовместимых сложных событий: 1) полярность для $\hat{x}_{\text{к}j}$ воспроизведена на выходе фазового детектора правильно, а для $\hat{x}_{\text{к}(j-1)}$ ошибочно; 2) полярность для $\hat{x}_{\text{к}j}$ воспроизведена ошибочно, а для $\hat{x}_{\text{к}(j-1)}$ верно.

Вероятность каждого из таких сложных событий равна

$$p=p_1(1-p_1),$$

где p_1 — вероятность неправильного воспроизведения полярности отдельной посылки. Эта вероятность соответствует вероятности ошибки при когерентном приеме

противофазного ФМн сигнала и определяется выражением (3.105).

Учитывая парность ошибок, а также влияние шумов по тракту опорного напряжения, выражение для вероятности ошибки $p_{сп}$ при приеме по методу сравнения полярностей можно записать в виде

$$p_{сп} = 2p_{ФМ}(1 - p_{ФМ}) + p_{оп}, \qquad (4.55)$$

где $p_{оп}$ — вероятность ошибки из-за действия шума в опорном тракте.

Как правило, в опорном тракте удается обеспечить более высокую степень фильтрации помех, чем в основном. Это объясняется тем, что полоса пропускания в опорном тракте может быть взята значительно уже полосы основного тракта. В таком случае влиянием ошибок из-за скачков фазы опорного напряжения под действием помех можно пренебречь. Поэтому для оценки помехоустойчивости когерентного приема сигналов с ОФМн можно пользоваться первым слагаемым выражениям (4.55), которое с учетом (3.105) принимает вид

$$p_{сп} \approx 2[0{,}5 - \Phi(\sqrt{2}h)][0{,}5 + \Phi(\sqrt{2}h)] =$$
$$= 2[0{,}25 - \Phi^2(\sqrt{2}h)]. \qquad (4.56)$$

При достаточно высоких требованиях к качеству передачи ($p_{сп} \ll 1$) выражение (4.56) можно записать в виде

$$p_{сп} \approx 2[0{,}5 - \Phi(\sqrt{2}h)] \qquad (4.57)$$

или, используя разложение (3.102), в виде

$$p_{сп} \approx \frac{1}{\sqrt{\pi}h}\exp(-h^2). \qquad (4.58)$$

Когерентный прием ОФМн сигналов по способу сравнения полярностей или с перекодированием можно вести в каналах не только с постоянными параметрами, но и с медленными случайными изменениями фазы принимаемых сигналов. Такие изменения могут «отслеживаться» системой ФАП в тракте формирования опорного напряжения. Если изменения параметров канала приводят к сравнительно быстрым флюктуациям фазы принимаемых сигналов, опорное напряжение перестает достаточно точно следить за фазой сигнала и условие

когерентности приема нарушается. Это приводит к значительному возрастанию вероятности ошибки приема, что связано с увеличением вероятности перескоков фазы опорного напряжения и ухудшением энергетического отношения сигнал/шум при нарушении когерентности.

4.7.2. Автокорреляционный (некогерентный) метод приема

Этот метод приема основан на том, что в качестве опорного напряжения используется предыдущая посылка сигнала, которая запоминается на время, равное ее длительности. В фазовом детекторе происходит сравнение фазы принятой посылки с фазой предыдущей (опорной) посылки.

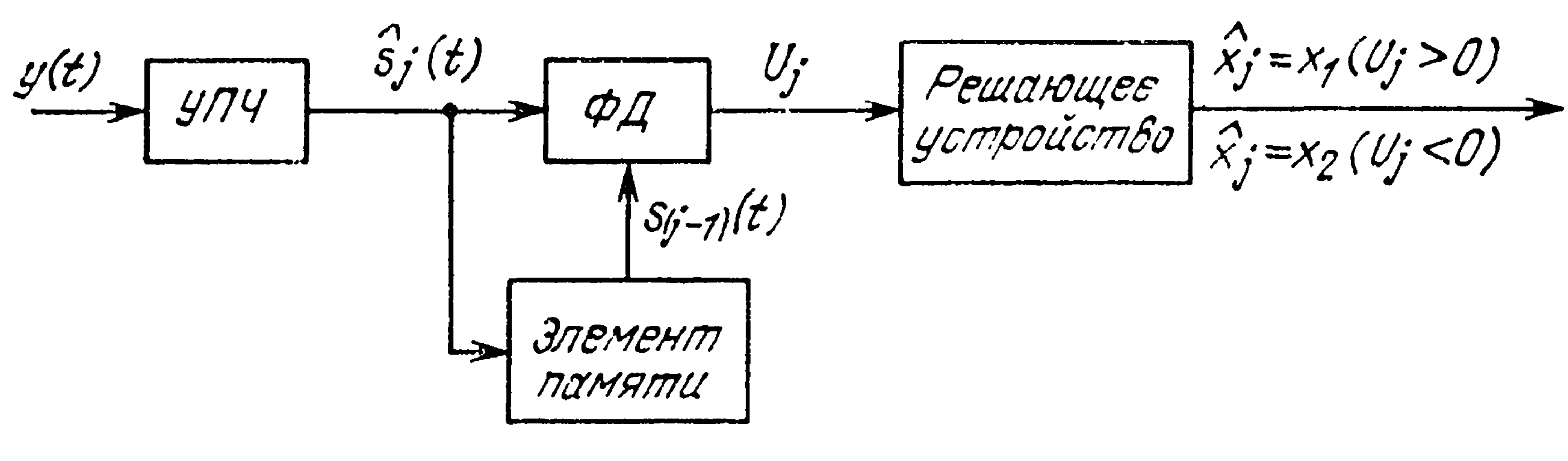


Рис. 4.16.

Структурная схема демодулятора ОФМн сигналов в этом случае имеет вид, показанный на рис. 4.16. Решение о том, какой символ был передан, принимается в соответствии с полярностью напряжения U_j на выходе фазового детектора: при положительном напряжении решающее устройство вырабатывает символ 1, а при отрицательном — символ 0. При таком способе приема фазовый детектор, по сути дела, вычисляет коэффициент автокорреляции сигнала на интервале, равном длительности двух посылок. Поэтому такой способ приема является автокорреляционным. В отечественной литературе обычно его называют способом сравнения фаз посылок [11]. В иностранной литературе этот способ получил название дифференциально-когерентного (разностно-когерентного) *.

* Differentialy coherent system — дифференциально-когерентная система.

Дальше будем называть этот способ «сравнением фаз», поскольку это название хорошо отражает суть дела и получило широкое распространение в литературе и инженерной практике.

Особенности приема ОФМн сигналов способом сравнения фаз иллюстрируются табл. 4.4. Для простоты рассмотрения действие помех не учитывается.

Таблица 4.4

Номер столбца j	0	1	2	3	4	5	6	7	8	9	10	11	12
x_j	1	0	1	0	0	1	1	0	0	0	1	1	1
s_j ОФМн	s_1	s_2	s_2	s_1	s_2	s_2	s_2	s_1	s_2	s_1	s_1	s_1	s_1
$s_{оп}=s_{(j-1)}$		s_1	s_2	s_2	s_1	s_2	s_2	s_2	s_1	s_2	s_1	s_1	s_1
полярность U_j		—	+	—	—	+	+	—	—	—	+	+	+
$\hat{x}_j$		0	1	0	0	1	1	0	0	0	1	1	1

Здесь приняты те же обозначения, что и в табл. 4.3. Нетрудно видеть, что прием ОФМн сигналов способом сравнения фаз принципиально исключает возможность обратной работы фазового детектора.

При приеме способом сравнения фаз опорным напряжением для поступившего на вход фазового детектора колебания

$$y_j(t) = s_j(t) + n_j(t)$$

является задержанное на время, равное длительности посылки, колебание $y_{j-1}(t)$, т. е.

$$y_{j-1}(t-\tau_0) = s_{j-1}(t-\tau_0) + n_{j-1}(t-\tau_0).$$

Если бы шумы отсутствовали, то разность фаз колебаний y_j и y_{j-1} полностью определялась бы только исходной информационной последовательностью символов, а также возможным изменением фазы каждой посылки при изменяющихся условиях распространения. Однако для соседних посылок эти изменения, как правило, настолько незначительны, что ими можно пренебречь *.

* Часто возможные случайные изменения параметров канала происходят медленно по сравнению с длительностью посылки. Подробнее об этом говорится в следующей главе.

Под действием шума фазы колебаний y_j и y_{j-1} приобретают случайный характер. Очевидно, что напряжение на выходе фазового детектора при сравнении фаз «зашумленных» посылок будет иметь положительную полярность, если фазы этих «зашумленных» посылок отличаются не более чем на $\pm\pi/2$, и отрицательную в противном случае.

Существует определенная вероятность того, что под действием помехи фаза посылки может измениться весьма существенно и в результате на выходе фазового детектора появится напряжение с полярностью, обратной той, которая должна была бы быть в отсутствие помех. Это приведет к появлению ошибочного символа на выходе решающего устройства. Поскольку каждая посылка участвует в формировании напряжения на выходе фазового детектора дважды (первый раз как посылка, а второй — как опора), то ошибки на выходе будут появляться, как правило, парами (сдвоенные ошибки).

Анализ вероятности ошибки при приеме ОФМн сигналов сравнением фаз выполнен в работах [2, 7, 11] и дает следующий результат:

$$p_{\text{сф}}=0{,}5\exp(-h^2), \qquad (4.59)$$

где h^2 — отношение мощности сигнала к мощности шума на входе приемника, определяемое выражением (4.32).

Формула (4.59) может быть получена различными путями, но все они исходят из того, что шумы в соседних посылках некоррелированы, а рассматриваемый способ приема некогерентен, поскольку фаза опорной посылки является случайной по отношению к фазе принимаемой посылки.

Как указывалось выше, для приема ОФМн сигналов способом сравнения фаз необходимо обеспечивать запоминание принимаемой посылки на время, равное ее длительности, для того, чтобы использовать ее в качестве опорного напряжения при приеме следующей посылки. По сути дела, такое запоминание необходимо лишь для того, чтобы сохранить фазу посылки. При этом вовсе не обязательно точно воспроизводить форму огибающей самой посылки.

Одна из наиболее применяемых схем демодулятора при приеме ОФМн сигналов способом сравнения фаз представлена на рис. 4.17. Основными элементами схе-

мы являются коммутируемые высокодобротные полосовые кинематические фильтры *Ф1* и *Ф2*, настроенные на центральную частоту спектра посылок. Работа схемы иллюстрируется эпюрами рис. 4.18. При появлении посылки s_{j-1} в момент t_1, ключ *К1* подключает ее на время τ_0 к фильтру *Ф1*. Для того чтобы обеспечить в этом фильтре начальные нулевые условия (погасить колебания, которые имелись в фильтре *Ф1* до появления посылки s_{j-1}), выход фильтра *Ф1* в момент t_1 на короткое

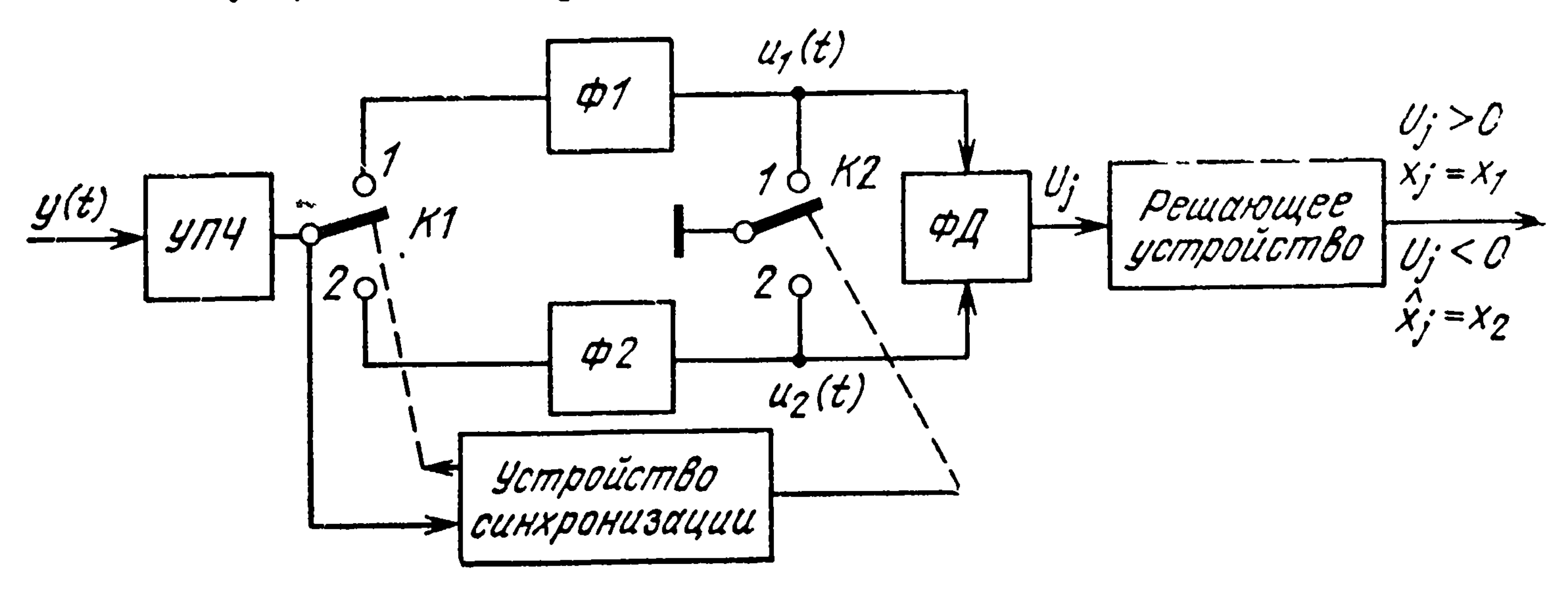


Рис. 4.17.

время ($\Delta\tau_0 \ll \tau_0$) подключается ключом *К2* к точке общего потенциала схемы (к «земле»). После окончания посылки в фильтре *Ф1*, сохраняются колебания с фазой посылки s_{j-1}. Эти колебания поступают на фазовый детектор и выполняют роль опорного напряжения для посылки s_j. В момент t_2, соответствующий появлению посылки s_j, ключи *К1* и *К2* переключаются в положение *2*, соответственно на время τ_0 и $\Delta\tau_0$ и в фильтре *Ф2* протекают процессы, аналогичные процессам в фильтре *Ф1*. Нетрудно видеть, что все посылки с нечетными номерами возбуждают один фильтр, а с четными — другой.

Таким образом, на выходе фильтров образуются колебания $u_1(t)$ и $u_2(t)$, которые сдвинуты друг относительно друга на интервал времени τ_0. Каждое из этих колебаний имеет длительность, равную $2\tau_0$. При этом первая часть колебания длительностью τ_0 соответствует принятой посылке, а вторая, с той же длительностью,— опорному напряжению для следующей посылки. Напряжение на выходе фазового детектора имеет вид линейно-нарастающих импульсов, полярность которых определяется значением разности фаз сравниваемых посылок. Если разность фаз посылок не превышает значения $\pm\pi/2$, импульсы имеют положительную полярность,

а если превышает — отрицательную. В соответствии со знаком полярности этих импульсов решающее устройство вырабатывает тот или иной информационный символ. Так как решающее устройство реагирует не только на полярность, но и на абсолютную величину импульсов, которая под действием помехи принимает случайные значения, возможны и ошибочные решения. Вероятность таких ошибок определяется формулой (4.59). Синфаз-

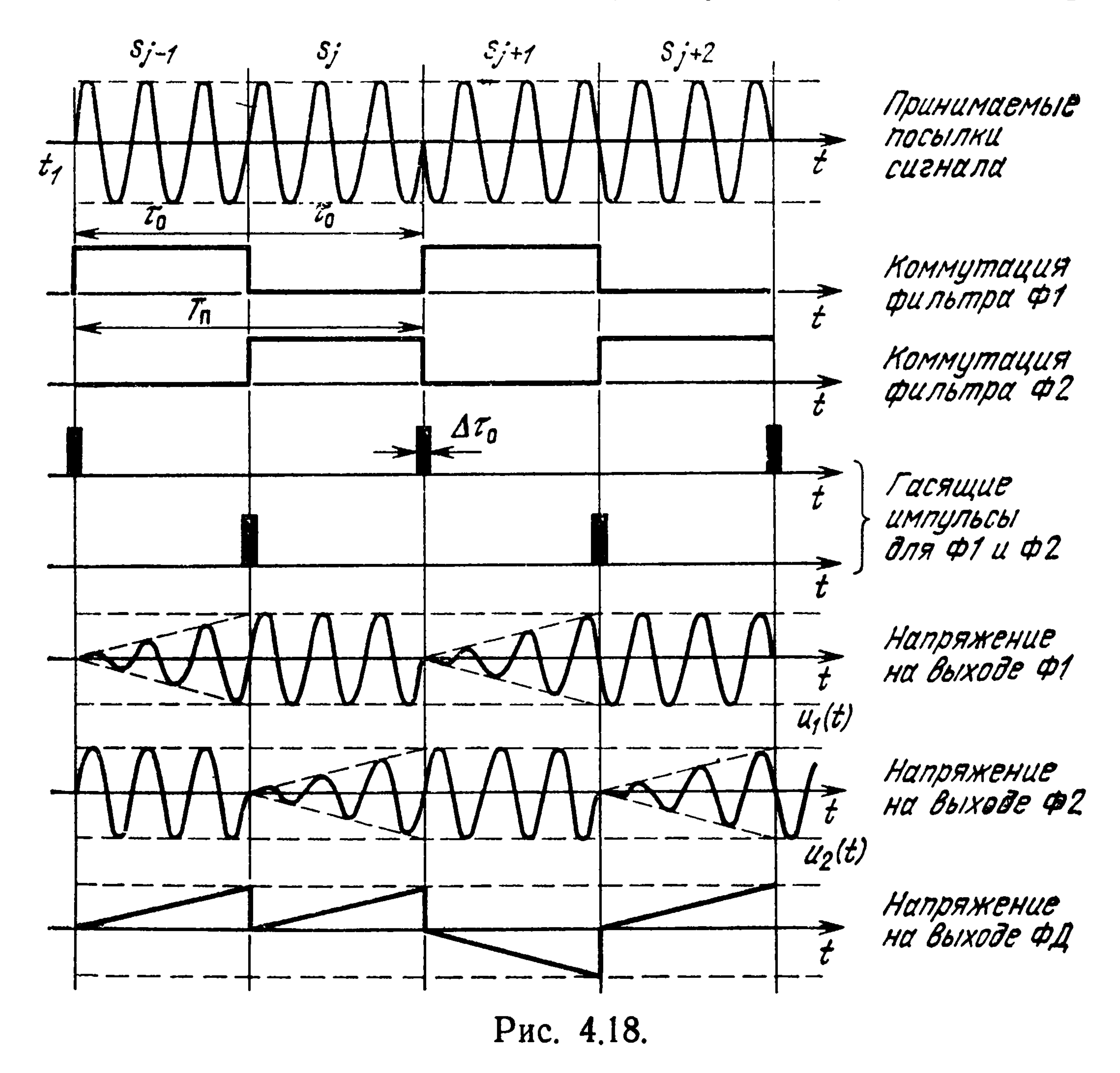


Рис. 4.18.

ное с границами посылок включение и выключение фильтров, а также принудительное гашение свободных колебаний в фильтрах обеспечивается специальными электронными устройствами коммутации, управляемыми с помощью отдельного канала синхронизации.

Рассмотренную схему демодуляции ОФМн сигналов называют схемой с коммутируемыми или кинематическими фильтрами. В ряде работ такую схему называют также схемой с синхронными фильтрами. Необходимо иметь в виду, что при использовании кинематических фильтров нужно учитывать ряд ограничений технического характера.

Рассмотрим суть этих ограничений. Для уменьшения потерь энергии посылки при демодуляции необходимо, чтобы время гашения колебаний $\Delta\tau_0$ в фильтре составляло малую часть длительности посылки τ_0. При допустимых потерях энергии не более 10% (до 0,5 дБ) необходимо обеспечить выполнение условия

$$\Delta\tau_0 \leqslant 0{,}1\,\tau_0. \qquad (4.60)$$

Минимальное время гашения колебаний в фильтре занимает приблизительно от 5 до 10 периодов, соответствующих центральной частоте настройки фильтра f_0:

$$\Delta\tau_0 \approx (5-10)/f_0. \qquad (4.61)$$

Чтобы фильтр был близок к идеальному интегратору, т. е. обеспечивал линейное возрастание огибающей выходного напряжения на интервале времени $(0-\tau_0)$, следует выполнить условие

$$\Delta f_ф \tau_0 \leqslant 0{,}1. \qquad (4.62)$$

Из условий (4.60)—(4.62) следует, что необходимая добротность фильтра $Q = \Delta f_ф/f_0 \approx 500-1000$. Проектирование фильтров с такими добротностями на частотах выше 10—20 кГц связано с большими трудностями. В настоящее время кинематические фильтры создают на частоты не выше 5—10 кГц.

Условия (4.60) и (4.61) устанавливают еще одно ограничение

$$R_т = \frac{1}{\tau_0} \leqslant \frac{0{,}1}{(5-10)} f_0 \approx (1 \div 2)\cdot 10^{-2} f_0. \qquad (4.63)$$

Это ограничение означает, что скорость манипуляции (техническая скорость передачи в бодах) должна быть примерно на два порядка ниже центральной частоты фильтра. Если принять $f_0 \approx 5-10$ кГц, то $R_т \approx 50-$ -200 бод.

С учетом отмеченных особенностей применение ОФМн сигналов при приеме с использованием кинематических фильтров наиболее целесообразно в каналах с невысокими скоростями передачи. Очень широкое применение рассмотренный способ приема находит при передаче данных и другой цифровой информации по телефонным и телеграфным каналам. В таких каналах обычно применяется уплотнение по частоте, чтобы обеспечить на

каждой поднесущей выполнение отмеченных выше условий.

Заметим, что прием способом сравнения фаз можно осуществлять и для перекодированной последовательности ОФМн сигналов. В этом случае схему рис. 4.16 нужно дополнить устройством, которое должно инвертировать выходную последовательность символов (переводить символ 1 в символ 0, а символ 0 в 1).

Если требуется обеспечивать передачу информации с высокими скоростями (десятки и сотни килобит), необходимо переходить к другим способам запоминания посылок (например, с помощью линий задержки) или использовать когерентные способы приема ОФМн сигналов.

4.7.3. Сравнение корреляционного и автокорреляционного методов приема ОФМн сигналов

Из приведенного выше рассмотрения следует, что при приеме ОФМн сигналов как при корреляционном, так и автокорреляционном методах используется опорное напряжение, формируемое из принимаемого сигнала, искаженного шумами. Однако степень фильтрации при формировании опорного напряжения в рассмотренных методах различна. При автокорреляционном методе фильтрация помех осуществляется только на интервале, равном длительности посылки. При корреляционном же методе фильтрация помех в тракте опорного напряжения может выполняться на значительно большем интервале. В пределе, если параметры канала постоянны, можно обеспечить сколь угодно высокую степень фильтрации помех. Поэтому формулы (4.58) и (4.59) можно рассматривать как нижнюю и верхнюю границы вероятности ошибки приема сигналов ОФМн. Нижняя граница относится к идеальному опорному напряжению, а верхняя — к «зашумленному» опорному напряжению.

Проведем сравнение этих границ при условии, что требования к качеству передачи довольно высоки ($p_{\text{ош}}<10^{-3}$). В этом случае вероятность ошибки $p_{\text{ош н}}$, соответствующая нижней границе, определяется выражением (4.58) и может быть записана в виде

$$p_{\text{ош н}}=\frac{1}{\sqrt{\pi}\,h_{\text{н}}}\exp(-h^2_{\text{н}}), \qquad (4.64)$$

где $h_н$ — отношение сигнал/шум для нижней границы.

Вероятность ошибки $p_{ош\,в}$, соответствующая верхней границе с учетом выражения (4.59), может быть записана в виде

$$p_{ош\,в}=0{,}5\exp(-h^2_в), \qquad (4.65)$$

где $h_в$ — отношение сигнал/шум для верхней границы.

Сравнивая (4.64) и (4.65) при одинаковых вероятностях ошибок, нетрудно получить выражение, характеризующее различие в энергетических затратах:

$$\gamma_h=\frac{h^2_в}{h^2_н}=1+\frac{1}{h^2_н}\ln\frac{\sqrt{\pi}\,h_н}{2}\approx 1+\frac{\ln h_н}{h^2_д}. \qquad (4.66)$$

При $h_н>3$ это различие становится незначительным (менее 10%). Сравнение (4.64) и (4.65) при одинаковых энергетических затратах дает следующий результат:

$$\gamma_p=p_{ош\,в}/p_{ош\,н}=\sqrt{\pi}\,h_н/2\approx 0{,}9h_н. \qquad (4.67)$$

Результаты сравнения показывают, что при достаточно высоких требованиях к точности передачи информации различие в помехоустойчивости корреляционного метода приема с идеальным (не «зашумленным») опорным напряжением и автокорреляционного метода с «зашумленным» опорным напряжением невелико. С учетом реальных условий формирования опорного напряжения при когерентном приеме, когда необходимо обеспечивать «отслеживание» медленных случайных изменений фазы сигнала, это различие будет еще меньше.

Основное различие между корреляционным и автокорреляционным методом приема сигналов ОФМн определяется не помехоустойчивостью, а скоростью передачи, которая может быть достигнута при реализации этих методов. Как указывалось выше, при автокорреляционном приеме трудно получить высокую скорость передачи. Для корреляционного приема эти трудности значительно меньше. Указанные особенности определили области применения обоих методов приема ОФМн. Автокорреляционный прием находит широкое применение в системах передачи данных по стандартным телефонным каналам, а корреляционный — в таких системах передачи цифровой информации, где требуются сравнительно высокие скорости передачи (радиотелеметрия, передача изображений и т. п.).

4.8. ВЛИЯНИЕ МЕЖСИМВОЛЬНЫХ ИСКАЖЕНИЙ И НЕИДЕАЛЬНОСТИ СИНХРОНИЗАЦИИ НА ПЕРЕДАЧУ ЦИФРОВОЙ ИНФОРМАЦИИ

В реальных системах связи помимо неизбежного действия случайных аддитивных помех существенное влияние на качество передачи цифровой информации могут оказывать межсимвольные искажения и неидеальность синхронизации. Рассмотрим кратко суть этих явлений.

4.8.1. Межсимвольные искажения

Прохождение информационной последовательности символов и соответствующих им сигналов через отдельные звенья системы связи (выходные каскады передатчика, избирательная часть приемника) сопровождается переходными процессами, обусловленными конечной полосой пропускания этих звеньев.

Переходные процессы приводят не только к некоторому искажению огибающей каждой посылки сигнала, но и наложению на нее остаточных напряжений от предшествующих посылок. Действие этих остаточных напряжений эквивалентно действию некоторой дополнительной помехи, которую принято называть межсимвольной, или интерференционной. Каналы, для которых характерны такие помехи, называются каналами с памятью. Заметим, что причиной межсимвольных помех могут быть также случайные изменения характеристик среды и многолучевое распространение сигналов (см. гл. 5).

Межсимвольные помехи представляют собой случайный процесс, характеристики которого определяются характеристиками случайной последовательности информационных символов и видом общей импульсной переходной характеристики канала связи (с учетом избирательных звеньев системы связи). В частности, если импульсная переходная характеристика монотонна, то закон распределения межсимвольных помех приближается к гауссовскому; если же характеристика немонотонна (имеет колебательный характер), то закон распределения существенно отличается от гауссовского.

Строгая теория помехоустойчивости систем передачи цифровой информации с учетом аддитивных шумов и межсимвольных помех до сих пор не разработана. Это

объясняется рядом причин: трудностью получения точного выражения функции плотности вероятности межсимвольных помех; разнообразием исходных данных при анализе; неаддитивностью межсимвольных помех и неравномерностью их энергетического спектра.

В настоящее время известен ряд приближенных методов для оценки помехоустойчивости приема с учетом межсимвольных помех [17—19]. Заметим, что даже при использовании приближенных методов расчеты помехоустойчивости, как правило, приходится выполнять на ЭВМ.

Наиболее простой метод приближенного учета межсимвольных помех состоит в том, что для каждого информационного символа во внимание принимается мешающее действие только предыдущего символа, а действие всех остальных предшествующих символов не учитывается.

Если предположить, что частотная характеристика, обусловленная всеми полосовыми фильтрами системы связи, симметрична и расстройка между центральной частотой спектра посылки сигнала и частотой настройки фильтров отсутствует, то можно пользоваться низкочастотными эквивалентами [7].

Пусть импульсная характеристика низкочастотного эквивалента системы связи $h_{\text{н}}(t)$ известна, тогда огибающая посылки сигнала на выходе фильтров системы связи определится выражением

$$S_{\text{вых}}(t) = \int_0^\infty S_{\text{вх}}(\tau)\, h_{\text{н}}(t-\tau)\, d\tau, \tag{4.68}$$

где $S_{\text{вх}}(t)$ — огибающая посылки сигнала на входе системы.

Если эта огибающая прямоугольна:

$$S_{\text{вх}}(t) = \begin{cases} S_0 & 0 \leqslant t \leqslant \tau_0, \\ 0 & \text{для других } t, \end{cases}$$

то

$$S_{\text{вых}}(t) = S_0 \int_0^{\tau_0} h_{\text{н}}(t-\tau)\, d\tau. \tag{4.69}$$

Полная энергия посылки сигнала $E_с$ на выходе фильтров системы в этом случае равна

$$E_с = \frac{1}{2} \int_0^\infty S^2_{вых}(t)\, dt$$

или с учетом (4.69)

$$E_с = \frac{S^2_0}{2} \int_0^\infty \int_0^{\tau_0} \int_0^{\tau_0} h_н(t-\tau_1)\, h_н(t-\tau_2)\, d\tau_1 d\tau_2 dt. \qquad (4.70)$$

Учитывая, что при приеме используется только та часть энергии посылки сигнала на выходе фильтра, которая соответствует интервалу времени $[0, \tau_0]$, выражение (4.70) можно представить в виде двух составляющих:

$$E_с = \frac{1}{2} \int_0^{\tau_0} S^2_{вых}(t)\, dt + \frac{1}{2} \int_{\tau_0}^\infty S^2_{вых}(t)\, dt = \alpha_ф E_с + (1-\alpha_ф) E_с, \qquad (4.71)$$

где $\alpha_ф$ — коэффициент, учитывающий неоптимальную фильтрацию принятой посылки $(\alpha_ф < 1)$.

Первая составляющая выражений (4.71) является полезной, а вторая определяет энергию остаточного переходного процесса, который налагается на следующие посылки (символы), т. е. характеризует межсимвольную помеху. Следовательно, энергия, определяющая различение посылки сигнала в шуме с учетом межсимвольной помехи, может быть оценена разностью энергии посылки, учитываемой в интервале $[0, \tau_0]$, и энергии межсимвольной помехи:

$$E_{разл} = \alpha_ф E_с - (1-\alpha_ф) E_с = (2\alpha_ф - 1) E_с. \qquad (4.72)$$

При известной импульсной характеристике системы связи $h_н(t)$ и принятом способе манипуляции сигнала величину $E_{разл}$ можно рассчитать и, следовательно, оценить ухудшение помехоустойчивости за счет межсимвольных помех.

Одной из простейших мер уменьшения влияния межсимвольной помехи является расширение полосы пропускания фильтров. Однако при этом увеличивается действие внутренних шумов. Существует оптимальная поло-

са, при которой достигается минимальная вероятность ошибки с учетом шумов и межсимвольных помех. Исследования показывают [20], что эта полоса близка к оптимальной для шумов ($\Delta f_{опт}\tau_0 \approx 1$).

Расчеты, выполненные в ряде работ, показывают, что при оптимальной полосе действие межсимвольных помех эквивалентно снижению энергии посылки сигнала на 1—2 дБ по сравнению со случаем, когда эти помехи не учитываются [17—21].

Принципиальной особенностью межсимвольных помех является то, что их действие нельзя скомпенсировать увеличением энергии сигнала, поскольку при этом пропорционально увеличивается и энергия таких помех. Это означает, что при наличии межсимвольных помех увеличение энергии посылок сигнала повышает помехоустойчивость только до некоторого предельного значения, которое определяется отношением сигнала к межсимвольной помехе. Повышение помехоустойчивости в этом случае достигается компенсационными методами устранения основных составляющих межсимвольных помех. Такие методы основаны на применении трансверсальных (поперечных) фильтров, устройств авторегулировки схем символьной синхронизации и т. п.

С вопросами теории приема цифровой информации с учетом межсимвольных помех более детально можно ознакомиться в работах [17—23].

4.8.2. Неидеальность синхронизации

Для нормальной работы приемного устройства синхронной цифровой системы связи необходимо обеспечить строгое временное соответствие между принимаемой последовательностью посылок сигнала и режимом работы демодулятора и декодера приемника. Установление такого соответствия называется синхронизацией. Чтобы обеспечить синхронизацию приемника, нужно располагать сведениями о частоте сигнала, его начальной фазе, а также о моментах начала и окончания информационных символов, кодовых комбинаций и групп таких комбинаций. Эта совокупность сведений называется синхроинформацией. Чем точнее эти сведения, тем ближе к идеальному может быть осуществлен процесс синхронизации.

Процесс синхронизации можно разделить на пять этапов:

1) поиск и захват передаваемого сигнала;

2) измерение (оценка) частоты и начальной фазы принимаемого сигнала* (обеспечение высокочастотной синхронизации);

3) установление начала и окончания каждой элементарной посылки принимаемого сигнала (обеспечение символьной синхронизации); в литературе по цифровым системам символьную синхронизацию часто называют тактовой, понимая под тактом период повторения синхроимпульсов, равный длительности элементарного символа;

4) установление начала и окончания кодовой комбинации (обеспечение словной синхронизации);

5) установление начала или окончания кадра, содержащего несколько кодовых комбинаций от различных источников (обеспечение кадровой синхронизации).

Последние три этапа называют временной синхронизацией. В системах связи необходимости в решении задачи первого этапа часто не возникает, так как направление прихода сигнала к приемной антенне обычно известно заранее.

Синхроинформацию можно извлечь или непосредственно из информационного сигнала, или из синхросигнала, передаваемого по отдельному каналу. Такой канал может быть либо полностью автономным и создаваться на своей несущей частоте, либо формироваться при включении специальных символов синхронизации в информационную последовательность. На практике чаще предпочитают иметь отдельный канал синхронизации. Несмотря на то, что такой канал требует дополнительных затрат полосы и энергии, его применение оказывается целесообразным, так как эти затраты составляют сравнительно небольшую часть общих затрат, а система получается технически более удобной и простой.

Извлечение необходимой синхроинформации и управление в соответствии с ней работой демодулятора и декодера приемника осуществляется с помощью устройств синхронизации. Для высокочастотной синхронизации широкое применение находят устройства, основанные на методах автоподстройки частоты и фазы. Некоторые из них были рассмотрены в § 4.5.

---

* Для когерентного приема. Для некогерентного приема достаточно оценивать только частоту сигнала.

Временную синхронизацию обычно осуществляют автоматическим слежением за изменением временного положения синхросимволов, передаваемых по отдельному временному или частотному каналу. Такое слежение основывается на тех же принципах, что и слежение за частотой или фазой гармонического сигнала и реализуется различными следящими временными дискриминаторами.

Неидеальность работы синхронизирующих устройств проявляется в том, что между оцениваемым (измеряемым) параметром и его истинным значением всегда имеется некоторое рассогласование. Величина этого рассогласования зависит как от применяемого способа синхронизации, так и от отношения сигнал/шум в канале синхронизации. Анализ помехоустойчивости приема цифровой информации с учетом неидеальности синхронизации представляет собой довольно сложную самостоятельную задачу, которую мы рассматривать не будем.

Некоторое представление о степени влияния неидеальности синхронизации на помехоустойчивость приема цифровой информации можно получить из табл. 4.5, которая составлена на основе исследований, изложенных в работах [23—26].

Таблица 4.5

Вид синхронизации	Величина рассогласования	Требуемая вероятность ошибки	Необходимое увеличение энергии посылки
Высокочастотная (по частоте)	$\Delta f_p \leqslant 0,1/\tau_0$	10^{-4}—10^{-5}	1,15—1,25
Высокочастотная (по фазе)	$\Delta\varphi_p \leqslant (10—15)^\circ$	10^{-4}—10^{-5}	1,1—1,15
Временная (по времени прихода сигнала)	$\Delta\tau_p \leqslant 0,05\,\tau_0$	10^{-4}—10^{-5}	1,3—1,5

В этой таблице Δf_p, $\Delta\varphi_p$ и $\Delta\tau_p$ — среднеквадратичные значения рассогласования по частоте, фазе и времени прихода сигнала соответственно.

Таблица показывает, во сколько раз необходимо увеличить энергию посылки по сравнению с идеальной синхронизацией для получения той же вероятности ошибки.

Подчеркнем, что приведенные значения ориентировочны и для расчетов нужно пользоваться результатами указанных работ.

4.8.3. *Оценка помехоустойчивости и реальной скорости передачи двоичных сигналов*

Межсимвольные помехи и неидеальность синхронизации приводят к тому, что для получения требуемой точности передачи информации в реальных системах связи необходимо затрачивать бо́льшую энергию посылки сигнала, чем в системах, где эти факторы отсутствуют (идеализированные системы).

Нетрудно показать, что необходимая энергия посылки в реальной системе определяется выражением

$$E_р = E/\gamma_{мс}\gamma_{сх} = P_с\tau_0/\gamma_{мс}\gamma_{сх}, \qquad (4.73)$$

где E — энергия посылки сигнала с мощностью $P_с$ и длительностью τ_0 в системе с идеальной синхронизацией и без межсимвольных помех; $\gamma_{мс}$ и $\gamma_{сх}$ — коэффициенты, учитывающие влияние межсимвольных помех и неидеальность синхронизации, соответственно. Естественно, что $\gamma_{мс}<1$ и $\gamma_{сх}<1$. Чем меньше влияние рассматриваемых факторов, тем ближе коэффициенты $\gamma_{мс}$ и $\gamma_{сх}$ единице. Учитывая результаты предыдущего рассмотрения, ориентировочно можно считать, что в реальных системах

$$\gamma_{мс} \approx 0{,}7—0{,}9; \quad \gamma_{сх} \approx 0{,}6—0{,}8. \qquad (4.74)$$

Следовательно, в реальных системах для получения заданной вероятности ошибки передачи цифровой информации необходимо увеличивать энергию посылок в 1,5—2 раза по сравнению с идеализированными системами. Такое увеличение можно осуществить, либо повышая мощность сигнала, либо увеличивая длительность посылки. Несложно убедиться в том, что при одинаковой вероятности ошибки увеличение энергии посылки в реальной системе любым способом всегда эквивалентно увеличению длительности этой посылки по сравнению с длительностью посылки в идеализированной системе *.

* В идеализированной системе при увеличении мощности посылки и сохранении неизменной вероятности ошибки нужно соответственно уменьшать длительность посылки.

Приняв во внимание (4.73), нетрудно видеть, что указанное увеличение длительности посылки определяется выражением

$$\tau_{0р}/\tau_0 = 1/\gamma_{мс}\gamma_{сх}. \tag{4.75}$$

Из выражения (4.75) следует, что межсимвольные помехи и неидеальность синхронизации неизбежно приводят к снижению скорости передачи информации в реальных системах связи.

Из теории передачи сообщений известно, что информационная скорость передачи в симметричных двоичных каналах с постоянными параметрами определяется выражением

$$R = \frac{1}{\tau_0}[1 + p_{ош}\log_2 p_{ош} + (1 - p_{ош})\log_2(1 - p_{ош})]. \tag{4.76}$$

При увеличении отношения сигнал/шум в канале вероятность ошибки $p_{ош}$ уменьшается и скорость передачи стремится к величине

$$\lim_{p_{ош}\to 0} R = 1/\tau_0 = R_т, \tag{4.77}$$

где $R_т$ — техническая скорость передачи (см. гл. 1).

С достаточно хорошим приближением можно считать, что выражение (4.77) справедливо, если $p_{ош} < 10^{-2}$. С учетом выражений (4.75) и (4.77) снижение скорости передачи в реальных системах связи $R_р$ можно оценить формулой

$$R_р/R_т = \tau_0/\tau_{0р} = \gamma_{мс}\gamma_{сх}, \tag{4.78}$$

из которой с учетом (4.74) следует

$$R_р = 1/\tau_{0р} = (1/\tau_0)\gamma_{мс}\gamma_{сх}. \tag{4.79}$$

Проведенное рассмотрение показывает, что межсимвольные искажения и неидеальность синхронизации могут приводить к заметному уменьшению скорости передачи информации в реальных системах связи.

4.9. КРАТКИЕ ИТОГИ ГЛ. 4

1. В реальных условиях параметры принимаемого сигнала никогда не могут быть известны точно. Особенно чувствительным параметром сигнала является его

фаза. Поэтому даже незначительные случайные изменения параметров канала приводят к заметным случайным изменениям фазы.

2. Структура приемных устройств в каналах со случайными изменениями фазы может быть найдена на основе теории оптимального приема в таких каналах и некоторых инженерных соображений, учитывающих в первую очередь простоту реализации приемников.

3. Применяемые на практике реальные способы приема простых двоичных сигналов принято делить на когерентные и некогерентные. При когерентных способах необходимо измерять (оценивать) значение фазы принимаемого сигнала и использовать полученную информацию о фазе при демодуляции сигнала. Реализация когерентных способов связана со значительным усложнением приемного устройства, поэтому в ряде случаев на практике отдают предпочтение некогерентным способам приема.

4. Проведенные исследования показывают, что при достаточно высоких требованиях к достоверности передачи помехоустойчивость некогерентного приема простых сигналов незначительно отличается от помехоустойчивости оптимального приема таких сигналов в каналах с полностью постоянными параметрами.

5. Результаты рассмотрения позволяют записать общее выражение для вероятности ошибки при некогерентном приеме двоичных сигналов в виде

$$p_{\text{ош}} = 0{,}5 \exp(-\gamma^2_{\text{c}} h^2/2), \tag{4.80}$$

где γ_{c} — коэффициент, зависящий от вида и параметров манипуляции сигналов (см. гл. 3). Из этого выражения следует, что для получения заданной вероятности ошибки необходимо обеспечить отношение мощности сигнала к мощности шума на входе приемника, равное

$$h^2 \geqslant (P_{\text{c}}/P_{\text{ш}})_{\text{вх}} = (2/\gamma^2_{\text{c}}) \ln(1/2p_{\text{ош}}). \tag{4.81}$$

6. При использовании двоичных сигналов предпочтение нужно отдавать сигналам с активной паузой. Важным достоинством таких сигналов является то, что они обеспечивают работу приемника при нулевом пороге независимо от уровня принимаемого сигнала. Иными сло-

вами, режим работы приемника инвариантен * к уровню принимаемого сигнала.

7. Наиболее сложен с точки зрения реализации прием двоичных сигналов с фазовой манипуляцией. Трудности, обусловленные явлением обратной работы фазового детектора, не позволяют использовать для передачи цифровой информации классические противофазные сигналы.

Преимущества фазовой манипуляции в реальных каналах связи удается достаточно полно реализовать только методами относительной фазовой манипуляции.

8. Прием ОФМн сигналов может быть осуществлен как корреляционным (когерентным), так и автокорреляционным (некогерентным) методами. Для корреляционного метода необходимо располагать опорным напряжением, когерентным с принимаемым сигналом. Существует ряд способов, позволяющих успешно решать эту задачу. При автокорреляционном методе для каждой принятой посылки роль опорного напряжения выполняет предыдущая посылка, задерживаемая на время, равное длительности передаваемого символа.

Обеспечивая приблизительно одинаковую помехоустойчивость приема, указанные методы обладают различными возможностями по скорости передачи. При автокорреляционном методе приема возможная скорость передачи лежит в пределах 50—100 бод. Корреляционный метод приема позволяет получить значительно более высокие скорости передачи.

9. Передача цифровой информации методом ОФМн по каналам с медленно изменяющимися параметрами обладает свойствами инвариантности, т. е. прием сигналов ОФМн в таких каналах ведется так, будто канал имеет постоянные параметры.

10. Качественные показатели реальных систем передачи цифровой информации существенно зависят от межсимвольных искажений, возникающих в отдельных звеньях системы, а также от неидеальности работы устройств синхронизации. Эти факторы приводят как к снижению помехоустойчивости систем, так и к заметному уменьшению скорости передачи информации (до 1,5—2 раз, а иногда и более).

---

* Инвариантность (от латинского in — не, varians — изменяющийся) — неизменность свойств какой-либо системы при изменении некоторых параметров или свойств ее частей или возмущений, действующих на эту систему.

КОНТРОЛЬНЫЕ ВОПРОСЫ И ЗАДАЧИ

4.1. Указать основные особенности реальных условий приема двоичных сигналов в каналах с постоянными параметрами.

4.2. Какие трудности возникают при реализации оптимального приема в каналах с постоянными параметрами?

4.3. Какой смысл вкладывают в понятие «реальные способы приема»?

4.4. Объяснить, почему при некогерентном приеме простых двоичных АМн сигналов вероятность ошибки в основном определяется вероятностью ложного приема посылки? При каких условиях это утверждение станет неверным?

4.5. Зачем в приемнике АМн сигналов необходимо иметь автоматическую регулировку усиления? Почему такая регулировка не требуется в приемниках ЧМн и ФМн сигналов?

4.6. Найти удельные затраты полосы и энергии при передаче информации двоичными сигналами с АМн, если требуемая вероятность ошибки а) 10^{-4}, б) 10^{-5}, в) 10^{-6}.

4.7. Дать сравнительную характеристику применяемых способов приема простых ЧМн сигналов.

4.8. Найти удельные затраты полосы и энергии при передаче информации двоичными сигналами с ЧМн для тех же условий, что и в задаче 4.6.

4.9. Сравнить результаты решения задач 4.6 и 4.8 при следующих различных условиях: а) пиковая мощность сигнала одинакова; б) занимаемая полоса одинакова; в) длительность посылок одинакова.

4.10. В чем заключаются особенности приема простых ФМн сигналов?

4.11. Дать сравнительную характеристику различных схем формирования опорного напряжения.

4.12. В чем суть метода передачи информации ОФМн сигналами?

4.13. Передать двоичную последовательность 001100010110001 методом ОФМн с перекодированием и принять образованную последовательность сигналов методами сравнения фаз и сравнения полярностей.

4.14. Определить допустимое рассогласование между фазой опорного напряжения и фазой посылок при когерентном приеме противофазных ОФМн сигналов, если вероятность ошибки не должна увеличиваться более чем в n раз.

СПИСОК ЛИТЕРАТУРЫ

1. **Гуткин Л. С.** Теория оптимальных методов радиоприема при флюктуационных помехах. М., «Сов. радио», 1972.

2. **Теплов Н. Л.** Помехоустойчивость систем передачи дискретной информации. М., «Связь», 1964.

3. **Райс С.** Теория флюктуационных шумов. Пер. с англ. В кн.: «Теория передачи электрических сигналов при наличии помех». М., ИЛ, 1953.

4. **Левин Б. Р.** Теоретические основы статистической радиотехники. Т. 1 М., «Сов. радио», 1966.

5. **Тихонов В. И.** Статистическая радиотехника. М., «Сов. радио», 1966.
6. **Хворостенко Н. П.** Статистическая теория демодуляции дискретных сигналов. М., «Связь», 1968.
7. **Стейн С., Джонс Дж.** Принципы современной теории связи и их применение к передаче дискретных сообщений. М., «Связь», 1971.
8. **Борисов Ю. П., Пенин П. И.** Основы многоканальной передачи информации. М., «Связь», 1967.
9. **Котельников В. А.** Теория потенциальной помехоустойчивости. М., Госэнергоиздат, 1956.
10. **Градштейн И. С., Рыжик И. М.** Таблицы интегралов, сумм, рядов и произведений. М., Физматгиз, 1962.
11. **Петрович Н. Т.** Передача дискретной информации в каналах с фазовой манипуляцией. М., «Сов. радио», 1965.
12. **Фазовая** и относительная фазовая телеграфия. Сб. статей. М., «Связь», 1967.
13. **Шахгильдян В. В., Ляховкин А. А.** Системы фазовой автоподстройки частоты. М., «Связь», 1972.
14. **Заездный А. М., Окунев Ю. Б., Рахович Л. М.** Фазоразностная модуляция и ее применение для передачи дискретной информации. М., «Связь», 1967.
15. **Бухвинер В. Е.** Дискретные схемы в фазовых системах радиосвязи. М., «Связь», 1969.
16. **Витерби Э. Д.** Принципы когерентной связи. М., «Сов. радио», 1970.
17. **Малолепший Г. А.** Влияние избирательного тракта на помехоустойчивость некогерентного приема дискретных сигналов. «Электросвязь», 1970, № 7.
18. **Lugannani.** Intersimbol interference and probability of error in digital systems. — «IEEE Trans. on Information Theory», 1969, v. IT-15, № 6.
19. **Проукис Дж., Миллер Дж.** Адаптивный приемник для цифровой связи через каналы с интерференцией между символами. — «Зарубежная радиоэлектроника», 1970, № 2.
20. **Vay Jones J.** Filter distortion and intersimbol interference effects on PSK signal. — «IEEE Trans. on Communication Technology, 1971, c. COM. 19, N 12.
21. **Панкратов В. П.** Фазовые искажения и их компенсация. М., «Связь», 1974.
22. **Кловский Д. Д., Николаев Б. И.** Инженерная реализация радиотехнических схем. М., «Связь», 1975.
23. **Сервинский Е. Г.** Оптимизация систем передачи дискретной информации. М., «Связь», 1974.
24. **Свириденко С. С.** Основы синхронизации при приеме дискретных сигналов. М., «Связь», 1974.
25. **Гинзбург В. В., Каяцкас А. А.** Теория синхронизации демодуляторов. М., «Связь», 1974.
26. **Стиффлер Дж.** Теория синхронной связи. М., «Связь», 1975.

Глава 5

Прием двоичных сигналов в каналах со случайными параметрами

5.1. ВВЕДЕНИЕ

В двух предыдущих главах были рассмотрены вопросы приема двоичных сигналов в каналах с постоянными параметрами и аддитивным стационарным гауссовским шумом. К каналам подобного типа можно отнести ряд реальных каналов связи как со свободно распространяющимися сигналами (например, различные информационные каналы, использующие УКВ в пределах геометрической видимости, космические каналы связи), так и с канализируемыми или направляемыми сигналами (например, кабельные и волноводные каналы связи).

Наряду с такими каналами в настоящее время широко применяются каналы, параметры которых в процессе передачи информации непрерывно и случайным образом меняются. К ним относятся уже упоминавшиеся ранее тропосферные, ионосферные и метеорные каналы связи.

Класс используемых каналов со случайными параметрами в последние годы заметно расширился: появились каналы связи, использующие Луну как пассивный ретранслятор; исследуются и разрабатываются каналы связи с отражением от искусственно создаваемых в околоземном космическом пространстве дипольных поясов; ведутся интенсивные исследования гидроакустических каналов передачи информации [1—3].

Перечисленные виды каналов относятся к каналам со свободно распространяющимися сигналами. Дальнейшее рассмотрение будет проводиться только применительно к таким каналам *.

Каналы со случайными параметрами и свободно распространяющимися сигналами можно разделить на две

* К каналам со случайными параметрами можно отнести также некоторые каналы с направляемыми сигналами (например, коммутируемые каналы передачи данных с использованием проводных и кабельных линий). Такие каналы здесь не рассматриваются. Исследование подобных каналов представляет большое самостоятельное направление. Интересующихся отсылаем к специальной литературе [4].

группы: 1) каналы с прямой волной; 2) каналы с отраженно-рассеянной волной. В каналах первой группы существует геометрическая видимость между передатчиком и приемником и сигнал распространяется в условиях случайных изменений параметров среды. Примером подобных каналов являются наземные каналы оптической связи, гидроакустические каналы и т. п.

В каналах второй группы геометрическая видимость между передатчиком и приемником отсутствует и связь обеспечивается при использовании некоторых свойств неидеальных сред рассеивать и переизлучать энергию сигнала. Принимаемый сигнал в таких каналах является результатом переизлучения и рассеяния переданного сигнала некоторым объемом среды (рис. 5.1.). Каналы этой группы являются типичными каналами со случайными параметрами.

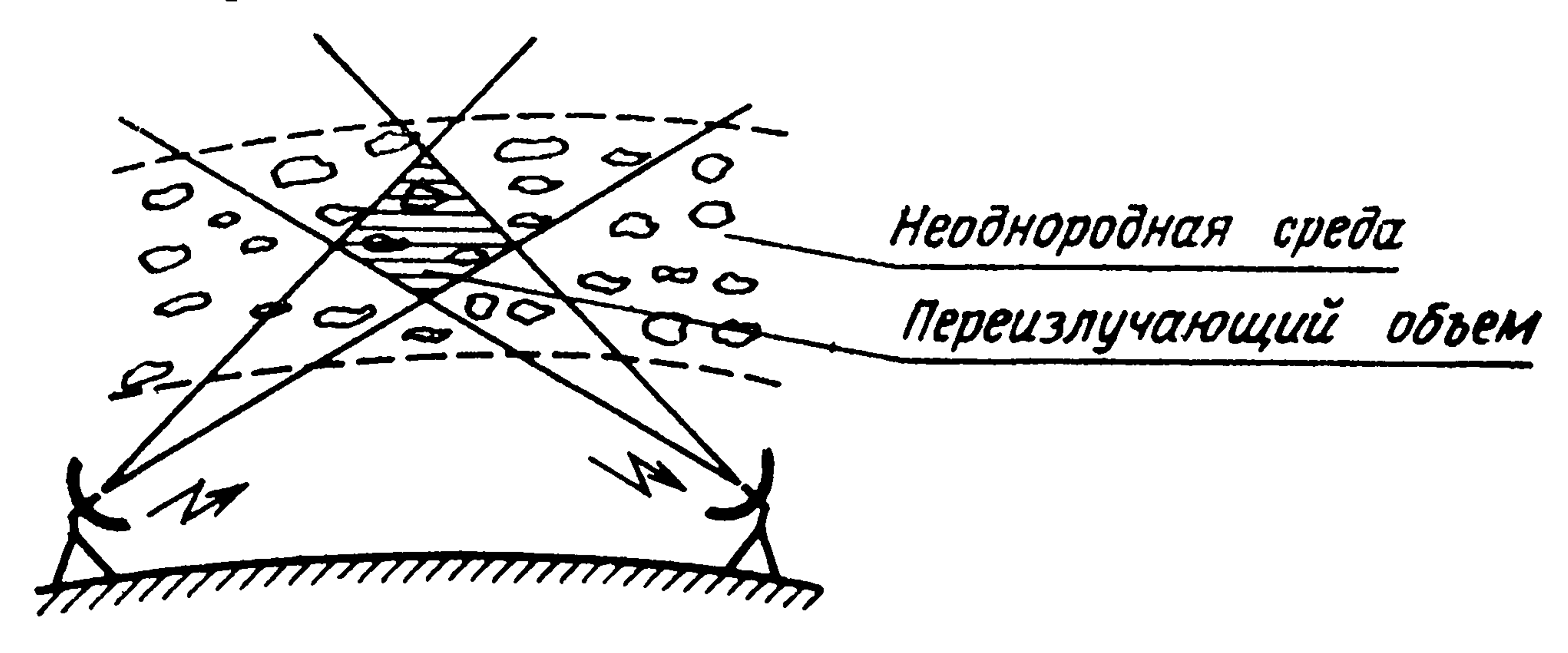


Рис. 5.1.

Каналы первой группы занимают промежуточное положение между каналами с постоянными параметрами и каналами второй группы. В зависимости от условий работы и параметров линии связи (состояния среды, дальности связи, характеристик антенн и т. п.) эти каналы иногда ближе к каналам с постоянными параметрами, а иногда — к типичным каналам со случайными параметрами.

Причиной случайного изменения параметров канала связи может быть действие целого ряда изменяющихся факторов (температуры, давления, плотности среды, наличия различных примесей и включений и т. п.). Под влиянием этих факторов в среде распространения происходят очень сложные физико-химические процессы, в результате которых ее структура и характеристики претерпевают непрерывные случайные изменения.

Для удобства исследований в теории распространения сигналов принято различать макро- и микроструктуру среды [6, 7].

Макроструктура отражает основные (глобальные) особенности строения среды и ее физические свойства, обусловленные некоторыми постоянно действующими факторами. Макроструктура среды и ее свойства меняются под действием метеорологических условий, а также суточных, сезонных и годовых изменений температуры. Изменения макроструктуры во времени протекают довольно медленно, и их средний период может составлять десятки минут и более.

Микроструктура отражает местные (локальные) особенности строения среды, обусловленные действием некоторых случайных факторов. К их числу относятся сложные вихревые (турбулентные) процессы, протекающие в среде, а также различные физико-химические превращения, происходящие в отдельных участках среды. Изменения микроструктуры могут происходить за время, составляющее доли секунды.

Изменения структуры среды и ее свойств оказывают существенное влияние на процесс распространения сигналов и приводят к случайным изменениям уровня сигнала, поступающего на вход приемника. При определенных условиях диапазон изменений уровня сигнала оказывается значительным, т. е. сигнал на входе испытывает глубокие замирания.

В механизме замирания сигналов различают две составляющие: одна из них обусловлена весьма медленными изменениями макроструктуры среды, а другая — сравнительно быстрыми изменениями микроструктуры.

Прием сигналов в каналах со случайными параметрами представляет собой значительно более сложную задачу, чем прием в каналах с постоянными параметрами, так как наряду с влиянием аддитивных помех на качество передачи информации существенное влияние оказывает явление замирания сигналов. Замирания приводят к тому, что на некоторых интервалах времени уровень принимаемого сигнала становится соизмеримым с уровнем аддитивных помех (или даже меньше этого уровня). В результате достоверность принятой информации резко ухудшается, а скорость передачи падает.

При исследовании вопросов приема информации в каналах со случайными параметрами полезна модель

среды распространения. Такая модель может быть построена на основе рассмотренной выше качественной картины распространения сигналов в каналах со случайными параметрами.

Несмотря на многообразие возможных сред распространения все они в первом приближении могут считаться линейными, т. е. сигнал в любой точке среды является линейной функцией возмущения (излученного сигнала). Это означает, что принятый сигнал $s_x(t)$ можно рассматривать как эффект на выходе некоторого линейного фильтра, на входе которого действует излученный сигнал $s_{\text{изл}}(t)$.

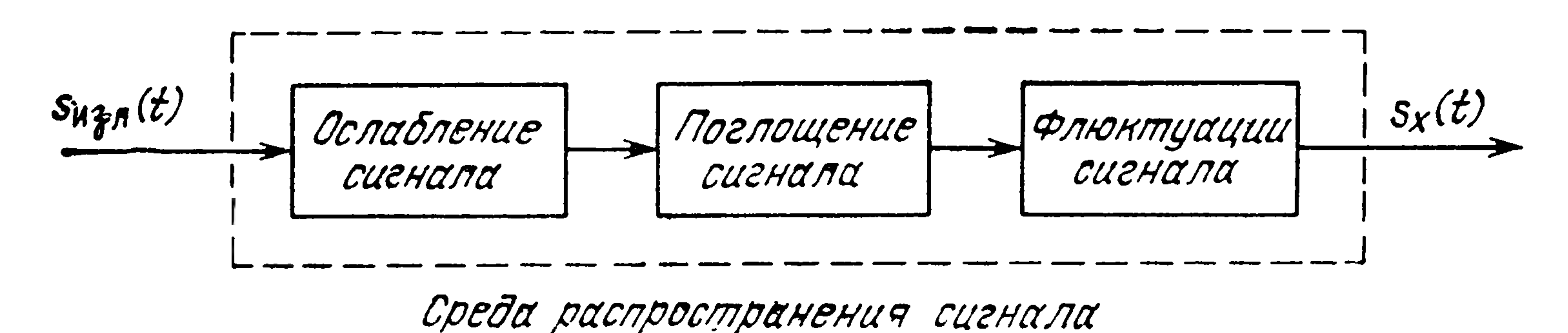


Рис. 5.2.

Чтобы линейная трактовка среды распространения сигналов была удобна для широкого класса каналов связи различной физической природы, целесообразно представить модель среды в виде трех последовательно соединенных пассивных линейных четырехполюсников (рис. 5.2), учитывающих разные стороны процесса распространения.

Первый четырехполюсник определяет ослабление сигнала при свободном распространении в идеальной среде. Его можно рассматривать как последовательное соединение ослабителя (аттенюатора) и линейного фильтра с равномерной частотной и линейной фазовой характеристиками на всех частотах. Ослабление сигнала в идеальной среде зависит только от геометрии распространения (дальности связи и характеристик приемной и передающей антенн). Величина этого ослабления может быть легко рассчитана по известной формуле дальности связи. Форма сигнала на выходе первого четырехполюсника остается такой же, как и на его входе, меняется только уровень этого сигнала.

Второй четырехполюсник отражает специфические особенности и физические свойства конкретной среды распространения. Он учитывает потери энергии сигнала в среде, связанные с поглощением, зависимость этих

потерь от частоты, а также медленные изменения структуры и свойств среды во времени и пространстве. Иными словами, второй четырехполюсник учитывает особенности макроструктуры среды и ее изменения. В результате этих изменений уровень сигнала на выходе второго четырехполюсника испытывает медленные замирания, которые наблюдаются на достаточно длительных интервалах времени (десятки минут и более). Эти замирания во многих случаях довольно хорошо можно аппроксимировать логарифмически нормальным законом

$$W(S_x)=\frac{1}{\sqrt{2\pi}\sigma(S_x-a)}\exp\left\{-\frac{[\log(S_x-a)-\overline{S}_x]^2}{2\sigma^2}\right\}, \quad (5.1)$$

$$W(S_x)=0, \quad S_x\leqslant a,$$

где S_x — случайное значение огибающей замирающего сигнала; $\overline{S}_x$ — среднее значение огибающей S_x; σ^2 — дисперсия; a — параметр, зависящий от свойств среды.

Необходимо отметить, что во многих каналах со случайными параметрами указанные замирания не приводят к значительным отклонениям огибающей сигнала от ее средних значений. Исключение составляют лишь некоторые каналы (например, каналы связи на коротких волнах, гидроакустические каналы значительной протяженности).

Характерной особенностью работы ряда систем связи в условиях медленных замираний сигнала является то, что за время сеанса связи состояние канала почти не меняется. Однако для разных сеансов связи это состояние может быть различным, меняясь от сеанса к сеансу случайным образом. Это означает, что медленные изменения макроструктуры среды делают условия работы линии связи нестационарными.

При фиксированной структуре системы связи и неизменных ее параметрах (ширине полосы пропускания, коэффициенте шума приемника, направленности антенн и т. п.) уменьшить вредное влияние медленных замираний можно увеличивая мощности излучаемого сигнала. При этом мощность передатчика проектируемой системы связи заранее выбирается с запасом, который учитывает возможность таких замираний (системный запас).

Трудности, возникающие из-за длительного ухудшения состояния канала, можно также преодолеть с помощью адаптивных систем, параметры которых меняют-

ся в процессе передачи в соответствии с изменениями условий в канале. Как правило, при уменьшении отношения сигнал/помеха процесс адаптации сводится к снижению скорости передачи увеличением длительности посылок сигнала и соответствующим уменьшением полосы пропускания приемника. Подобный метод адаптации находит широкое применение не только в каналах со случайными параметрами, но и в динамических гауссовых каналах, когда отношение сигнал/помеха падает при увеличении расстояния между передатчиком и приемником при движении объектов, образующих линию связи (характерным примером здесь является линия дальней космической связи).

Другим методом адаптации, позволяющим ослабить влияние длительных ухудшений состояния канала, является смена рабочей частоты в соответствии с прогнозом возможного состояния канала во время сеанса связи. Прогнозирование условий распространения сигналов осуществляется на основе накопленных априорных сведений о данных каналах. Такой метод находит широкое применение на линиях дальней коротковолновой радиосвязи, а также на тропосферных и ионосферных линиях связи.

Помимо медленных замираний, второй четырехполюсник должен учитывать зависимость поглощения энергии сигнала от частоты. В результате этой зависимости сигнал на выходе второго четырехполюсника будет искажаться по форме, поскольку его спектр отличается от спектра сигнала на входе. Таким образом, в общем случае второй четырехполюсник нужно рассматривать как фильтр со случайными и медленно меняющимися частотно-зависимыми характеристиками.

Для ряда каналов частотная зависимость характеристик среды проявляется незначительно, и в первом приближении с нею можно не считаться. К таким каналам можно отнести тропосферные (в них указанная зависимость практически не проявляется), ионосферные и метеорные. В некоторых каналах частотно-зависимые свойства среды могут быть заметно выраженными (например, в оптических наземных линиях связи, в гидроакустических каналах). Они имеют существенное значение при решении вопросов выбора формы сигнала способа приема и т. п.

Затронутые выше важные вопросы приема сигналов

в каналах с частотно-зависимыми характеристиками и медленно изменяющимися условиями распространения представляют отдельную большую проблему и выходят за рамки данного пособия.

Третий четырехполюсник в упрощенной модели учитывает сравнительно быстрые случайные изменения микроструктуры среды, которые приводят к соответствующим замираниям сигнала на входе приемника.

Трактовка среды распространения сигналов как линейного фильтра со случайными переменными параметрами в настоящее время широко распространена [1, 2, 5, 8], а представление среды в виде последовательного соединения четырехполюсников, учитывающих ее различные характерные особенности, является полезной моделью при исследовании различных вопросов передачи информации в каналах со случайными параметрами [9].

Основное внимание в этой главе уделено вопросам приема простых двоичных сигналов с учетом влияния микроструктуры среды и действия аддитивных помех типа гауссовских шумов. При этом полагаем, что за время сеанса связи макроструктура среды не изменяется, а параметры излучаемого сигнала выбраны так, чтобы с частотными свойствами среды можно было не считаться. Следовательно, при таких допущениях случайные изменения уровня сигнала, приходящего на вход приемника, и его возможные искажения обусловлены только случайными изменениями микроструктуры среды.

5.2. УПРОЩЕННАЯ МОДЕЛЬ ЗАМИРАНИЙ СИГНАЛА В СРЕДАХ СО СЛУЧАЙНО ИЗМЕНЯЮЩЕЙСЯ МИКРОСТРУКТУРОЙ

Сложные и непрерывные изменения микроструктуры среды приводят к появлению в ней неоднородностей. Различают неоднородности двух типов: слоистые и глобулярные.

Слоистые неоднородности возникают, если в среде появляются условия, при которых коэффициент преломления претерпевает скачок (в среде имеются протяженные участки с различными температурами или плотностями).

Глобулярные неоднородности являются результатом действия турбулентных процессов в среде. Под действием таких процессов в среде образуются локальные неоднородности (зерна неоднородности или глобулы), в пределах которых коэффициент преломления несколько отличается от значений в окружающей среде.

В общем случае форма возникающих неоднородностей произвольна, однако при построении упрощенных моделей среды часто принимают, что эти неоднородности имеют сферическую или эллипсоидальную форму. Такие неоднородности непрерывно возникают, перемещаются и исчезают. Интенсивность этого процесса, размер образующихся неоднородностей и их количество в единице объема зависят от характера и интенсивности турбулентных процессов в среде. Важно отметить, что подобные процессы существуют в любой среде и в ряде случаев могут оказывать заметное влияние на распространение сигналов.

Распространение сигнала в среде с неоднородностями сопровождается процессами отражения и рассеивания его энергии этими неоднородностями*. В результате излученный сигнал, пройдя через среду, попадает на вход приемника по многим путям (лучам). Это явление называется многопутевым распространением, или многолучевостью. Так как неоднородности среды, выполняющие роль элементарных «отражателей» или «рассеивателей» энергии сигнала, непрерывно и случайным образом меняются, то время прохождения сигнала от передатчика к приемнику по отдельным лучам различно и случайно. Это приводит к тому, что сигнал на входе приемника представляет собой сумму отдельных колебаний, время прихода и амплитуды которых случайны. Интерференция таких составляющих является основной причиной флюктуаций параметров сигнала, принимаемого в условиях многолучевости.

---

* Помимо отражения и рассеивания энергии сигнала неоднородностями, в самой среде происходит также поглощение его энергии. В некоторых средах оно может быть значительным. Характер указанных процессов и их интенсивность существенно зависят от свойств среды, природы сигнала (электромагнитное излучение, акустические колебания и т. п.), несущей частоты сигнала и ширины его спектра, размеров неоднородностей и т. п. Рассмотрение затронутых вопросов является предметом теории распространения сигналов в различных средах.

Рассмотрим подробнее некоторые особенности и свойства многолучевого сигнала. Пусть передана посылка сигнала

$$s_i(t) = Sf_i(t), \qquad (5.2)$$

где S — амплитуда посылки; индекс i относится к тому или иному передаваемому информационному символу x_i (в двоичных системах, например, $i=1$ соответствует передаче символа x_1, а $i=2$ — передаче символа x_2); $f_i(t)$ — посылка единичной амплитуды, соответствующая выбранному виду манипуляции сигнала и передаваемому символу x_i. Так например, для бинарной противофазной ФМн

$$f_i(t) = \begin{cases} \sin(\omega_0 t + \varphi_i), & 0 \leqslant t \leqslant \tau_0, \\ 0 & \text{для других } t, \end{cases} \qquad (5.3)$$

где τ_0 — длительность передаваемой посылки;

$$\varphi_i = \begin{cases} 0, & \text{если } x_i = x_1, \\ \pi, & \text{если } x_i = x_2. \end{cases}$$

Нетрудно записать выражение $f_i(t)$ и для любого другого вида манипуляции. Функция $f_i(t)$ отображает особенности конкретного вида манипуляции. В дальнейшем будем называть ее сигнальной функцией. В этой главе запись в виде (5.2) будет использоваться наиболее часто, так как она удобна при рассмотрении ряда вопросов. Такую запись можно рассматривать как частный случай представления сигнала с помощью комплексной огибающей (см. [7]).

Принятый сигнал можно записать в виде

$$s_x(t) = \sum_k s_{ik}(t) = \sum_k S_{ik}(t) f_i[t - \tau_{зk}(t)], \qquad (5.4)$$

где k — число лучей; $S_{ik}(t)$ — огибающая i-й посылки сигнала, принятого по k-му лучу; $\tau_{зk}$ — запаздывание составляющей k-го луча относительно момента начала излучения i-й посылки.

Динамический характер микроструктуры среды, проявляющийся в ее непрерывных случайных изменениях, приводит к случайным изменениям числа лучей, а также амплитуды и времени запаздывания каждой составляющей. Это, в свою очередь, вызывает случайные изменения многолучевого сигнала, которые проявляются в виде флюктуаций его амплитуды и времени запаздывания.

Таким образом, многолучевой сигнал необходимо рассматривать как некоторый случайный процесс.

Ряд таких важных параметров микроструктуры, как средний размер неоднородностей, среднее число неоднородностей в единице объема среды (плотность неоднородностей), их средняя скорость перемещения в среде и т. п., полностью определяется состоянием и параметрами макроструктуры, которые остаются практически неизменными в течение длительных интервалов времени (от нескольких минут до нескольких часов и более). Это дает основание считать, что на таких интервалах времени указанный случайный процесс стационарен.

Для описания стационарных случайных процессов, которые используются в качестве математических моделей замираний сигналов в средах со случайно изменяющейся микроструктурой, вводят ряд характеристик и определений. Рассмотрим их и выясним влияние некоторых характеристик на передачу цифровой информации.

5.2.1. Время многолучевого растяжения сигнала

Время запаздывания сигнала, пришедшего на вход приемника по лучу с номером k, можно представить в виде

$$\tau_{зk}=\tau_{з0}+\Delta\tau_{зk}, \tag{5.5}$$

где $\tau_{з0}$ — среднее по всем лучам время запаздывания; $\Delta\tau_{зk}$ — случайное отклонение времени запаздывания по k-му лучу от среднего времени запаздывания $\tau_{з0}$.

Очевидно, что для величины $\Delta\tau_{зk}$ можно записать условие

$$\min\Delta\tau_з\leqslant\Delta\tau_{зk}\leqslant\max\Delta\tau_з. \tag{5.6}$$

Временем многолучевого растяжения сигнала $\tau_р$ называют величину

$$\tau_р=\max\Delta\tau_з-\min\Delta\tau_з. \tag{5.7}$$

Она характеризует разницу между максимальным и минимальным временем запаздывания при многолучевом распространении сигнала,

Непрерывные изменения микроструктуры среды приводят к тому, что величина $\Delta\tau_{3k}$ может принимать любое значение внутри интервала многолучевого растяжения τ_p. В этом случае говорят, что имеет место непрерывная (дисперсная) многолучевость.

На распространение сигналов в средах со случайными параметрами существенное влияние могут также оказывать слоистые неоднородности и границы раздела сред с различающимися физическими свойствами (например, «тропосфера — ионосфера», «тропосфера — земная по-

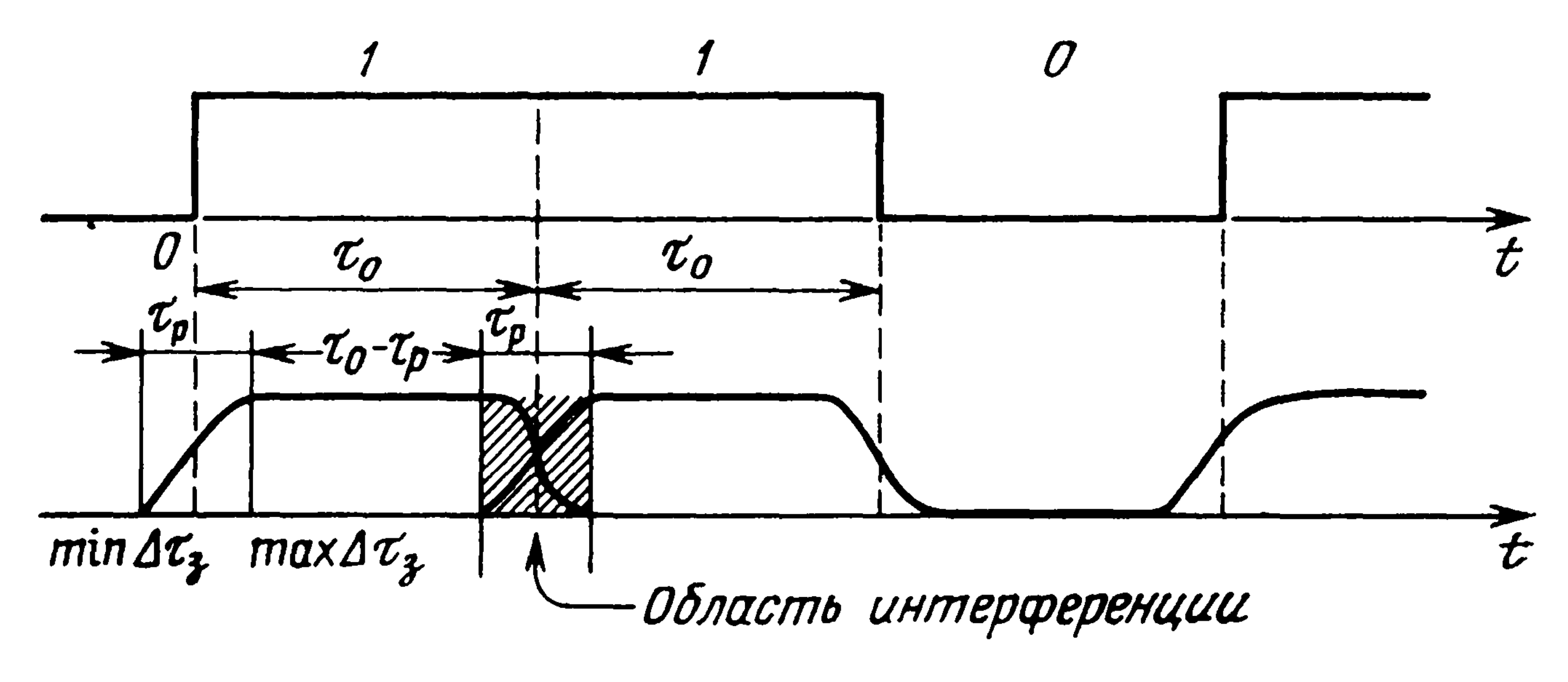


Рис. 5.3.

верхность», «море — атмосфера» и т. п.). Влияние указанных факторов характеризуется сравнительно медленными случайными изменениями во времени и приводит к тому, что запаздывания по некоторым лучам сосредоточены в некоторых неперекрывающихся участках. Таким образом, наряду с непрерывной многолучевостью может одновременно наблюдаться и дискретная многолучевость.

Эффект многолучевого растяжения приводит к тому, что каждой переданной посылке сигнала длительностью τ_0 на входе приемника соответствует посылка с длительностью, увеличенной на время растяжения τ_p. Такое растяжение длительности принимаемых посылок вызывает взаимное перекрытие принимаемых символов (межсимвольную интерференцию). Сказанное поясняется рис. 5.3, где показаны огибающие нескольких переданных и принятых символов. Для удобства сопоставления огибающих среднее время запаздывания принятых символов условно взято равным нулю.

Межсимвольная интерференция, появляющаяся при многолучевом распространении, может приводить к существенному ухудшению качественных показателей при

передаче цифровой информации. При использовании простых двоичных сигналов уменьшить влияние межсимвольных искажений можно, так увеличивая длительности посылки τ_0, чтобы обеспечивалось условие

$$\tau_0 \gg \tau_p. \qquad (5.8)$$

В этом случае уменьшение энергии посылки при исключении перекрывающихся участков незначительно*.

Из (5.8) вытекает очень важное условие:

$$R_т = 1/\tau_0 \ll 1/\tau_p,$$

которое означает, что эффект многолучевого распространения в каналах со случайными параметрами может существенно ограничить скорость передачи.

Время многолучевого растяжения сигнала зависит от ряда факторов (типа канала, протяженности линии связи, состояния среды, диаграмм направленности антенн и т. п.) и для разных линий может существенно различаться. Например, в линиях дальней коротковолновой радиосвязи протяженностью от 1500 до 4000 км это время составляет от 1 до 3 мс.

Для тропосферных и ионосферных линий связи характерно применение антенных устройств с большим коэффициентом направленного действия. Поэтому в таких линиях время многолучевого распространения значительно меньше, чем в коротковолновых линиях связи. В тропосферных линиях протяженностью от 600 до 1000 км это время составляет единицы микросекунд (до 5 мкс); в ионосферных линиях протяженностью от 1000 до 3000 км — несколько десятков микросекунд (до 100 мкс).

5.2.2. Доплеровское растяжение спектра сигнала

Непрерывные изменения микроструктуры среды, проявляющиеся в случайных перемещениях элементарных неоднородностей (отражателей и рассеивателей), приводят к случайному доплеровскому смещению частоты спектральных составляющих сигнала, проходящего через

* При использовании сложных сигналов возможны такие методы приема, которые позволяют устранить влияние межсимвольной интерференции без уменьшения энергии принятого сигнала. Эти вопросы рассмотрены в гл. 6.

среду. Так как неоднородности перемещаются с различными скоростями, а сигнал приходит на вход приемника по нескольким путям, то его спектр отличается от спектра переданного сигнала. Каждая спектральная составляющая принятого сигнала приобретает случайные изменения, что эквивалентно ее «размытию» («уширению») в некоторой области частот $\Delta f_д$. Величина $\Delta f_д$ называется допплеровским растяжением частоты. Если допплеровское растяжение примерно одинаково для всех составляющих спектра флюктуирующего сигнала, то и скорость их случайных изменений во времени примерно одинакова. Диапазон частот, в пределах которого соблюдается указанное условие, называется полосой когерентности среды, или полосой частот когерентного распространения сигнала. Когерентную полосу частот часто называют также полосой частот гладких замираний.

Особенностью гладких замираний является то, что замирания всех спектральных составляющих сигнала протекают во врмени когерентно (иногда говорят «дружно») и в результате форма посылки сигнала не испытывает сколь-нибудь заметных искажений *.

Для выполнения условия когерентности необходимо, чтобы величина допплеровского растяжения $\Delta f_д$ была значительно меньше ширины спектра сигнала Δf_c:

$$\Delta f_д \ll \Delta f_c. \qquad (5.9)$$

Если принять, что $\Delta f_c \approx 1/\tau_0$, то условие (5.9) можно записать в виде

$$\tau_0 \ll 1/\Delta f_д. \qquad (5.10)$$

В этом условии величина $1/\Delta f_д$ может рассматриваться как некоторый период изменения состояния микроструктуры среды.

Чем выше скорость перемещения элементарных неоднородностей, тем больше допплеровское растяжение и тем быстрее происходит изменение микроструктуры среды. Условие (5.10) означает, что за время, равное длительности посылки сигнала, эти изменения должны быть незначительными.

Невыполнение условия (5.9) или (5.10) приводит к тому, что скорость замираний спектральных состав-

* Иными словами, для гладких замираний характерна сильная корреляция (близкая к единице) замираний всех спектральных составляющих, в результате которой форма спектра замирающего сигнала остается без изменений, а меняется лишь его масштаб.

ляющих сигнала становится разной, т. е. между замираниями на разных частотах нет сильной корреляции. В результате этого появляются заметные случайные искажения спектра, а следовательно, и формы посылки сигнала. Такое явление называют селективными замираниями. Прием сигналов в условиях селективных замираний очень часто невозможен из-за глубоких замираний, сопровождаемых сильными случайными искажениями формы передаваемого сигнала. Поэтому обычно стремятся выбирать параметры сигнала так, чтобы иметь дело только с гладкими замираниями.

Для того чтобы обеспечить малое влияние многолучевого и допплеровского растяжений, необходимо одновременно выполнить условия (5.8) и (5.10), т. е. выбрать длительность посылки сигнала в соответствии с неравенствами

$$\tau_р \ll \tau_0 \ll 1/\Delta f_д \tag{5.11}$$

или

$$\tau_р \Delta f_д \ll \tau_0 \Delta f_д \ll 1, \tag{5.12}$$

отсюда следует условие

$$k_р = \tau_р \Delta f_д \ll 1. \tag{5.13}$$

Величина $k_р$ называется коэффициентом растяжения. Это очень важный параметр, характеризующий возможности передачи цифровой информации по каналам со случайными параметрами.

Для выполнения условия (5.11) необходимо, чтобы коэффициент растяжения был много меньше единицы. Эти условия для многих каналов выполняются. Однако в некоторых каналах, таких, например, как каналы с использованием отражений от дипольных поясов и гидроакустические, коэффициенты растяжения могут приближаться к единице и замирания создают серьезные трудности в приеме сигналов.

5.2.3. Глубина и скорость замираний огибающей сигнала

Непрерывные и случайные изменения микроструктуры среды приводят к тому, что амплитуда посылок сигнала на входе приемника становится случайной величиной. При выполнении условий (5.11) или (5.13) эта величина может считаться постоянной в пределах длительности отдельной посылки и медленно изменяющейся от

посылки к посылке. Для характеристики этих изменений вводят понятия глубины и скорости замирания флюктуирующего сигнала. Рассмотрим суть этих понятий.

Если передать достаточно протяженную последовательность посылок сигнала, то огибающая этой последовательности на входе приемника будет изменяться случайным образом (рис. 5.4). Внутренняя структура

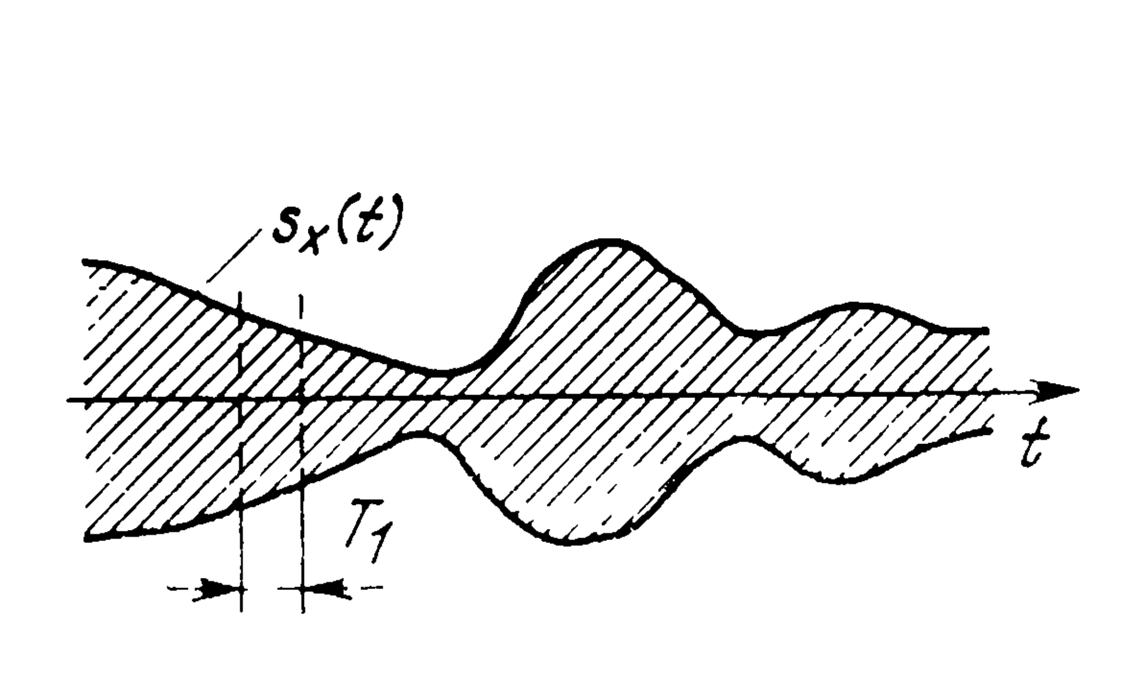


Рис. 5.4

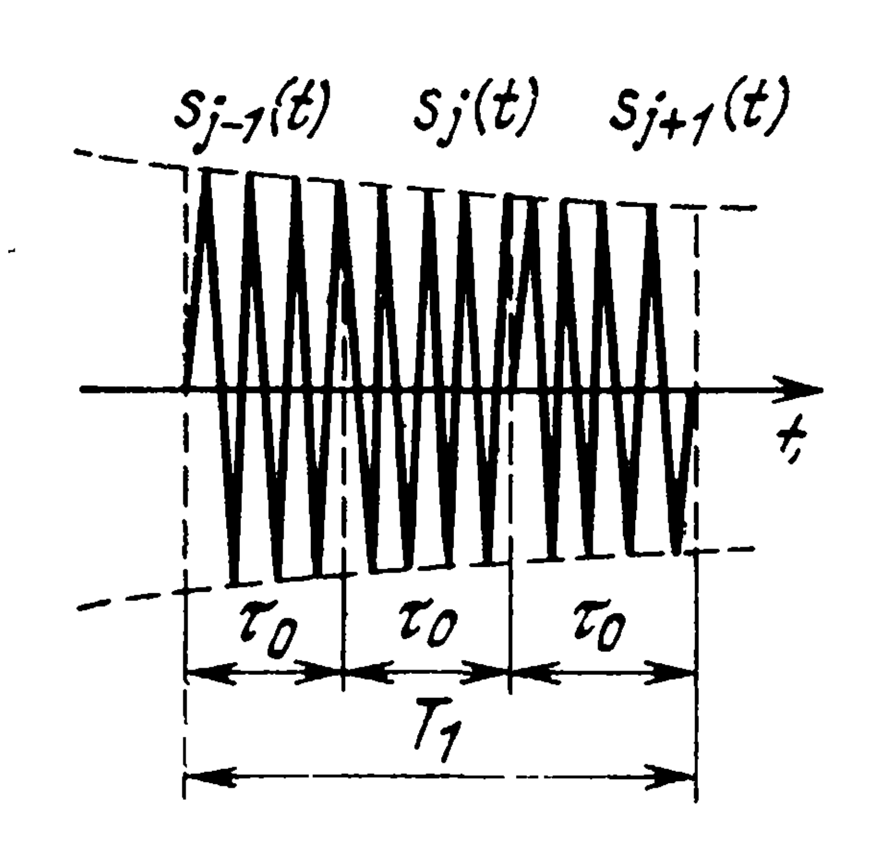


Рис. 5.5.

последовательности на рис. 5.4 для простоты не показана. В более крупном масштабе и в упрощенном виде (без интерференционных межсимвольных искажений) часть этой структуры на интервале T_1 приведена на рис. 5.5, где представлены три посылки сигнала s_{j-1}, s_j, s_{j+1}.

Глубину замираний принято характеризовать изменением уровня огибающей относительно ее медианного значения. Медианным называют такой уровень огибающей сигнала, при котором в течение сеанса связи (или интервала наблюдения при экспериментальных исследованиях свойств канала) суммарное время пребывания процесса, как выше, так и ниже этого уровня, одинаково.

Глубина замираний в линиях связи большой протяженности может быть значительной (от 20 до 30 дБ). К этому вопросу мы еще вернемся несколько позже.

Скорость замираний принято характеризовать средним числом односторонних пересечений (снизу вверх или сверху вниз) огибающей сигнала с медианным уровнем в секунду. Эту величину можно также трактовать как среднюю частоту замираний или как среднее число выбросов огибающей, превышающих медианный уровень в течение одной секунды.

Для измерения скорости замираний на практике обычно используют излучение непрерывной немодулированной несущей и измеряют частоту пересечения огибающей принимаемого флюктуирующего сигнала с ее медианным уровнем.

Если сделать запись флюктуирующей огибающей, то, обрабатывая ее, можно получить энергетический спектр или корреляционную функцию замираний. Величина скорости замираний однозначно связана с шириной энергетического спектра замираний или с их корреляционной функцией. Поэтому скорость замираний можно определить и по одной из этих характеристик. Поскольку эти характеристики полностью определяются физическими процессами в среде, они могут быть найдены только по экспериментальным данным.

Обычно при теоретическом рассмотрении каналов со случайными параметрами нормированную корреляционную функцию замираний, построенную по экспериментальным данным, аппроксимируют выражением

$$\rho(\tau) = \exp\left(-|\tau|/\tau_{\text{фл}}\right) \qquad (5.14)$$

или

$$\rho(\tau) = \exp\left(-\tau^2/2\tau^2_{\text{фл}}\right), \qquad (5.15)$$

где $\tau_{\text{фл}}$ — параметр, однозначно связанный со скоростью замираний.

Если время корреляции процесса определять в соответствии с выражением

$$\tau_{\text{к}} = \int\limits_0^\infty \rho(\tau)\, d\tau, \qquad (5.16)$$

которое широко применяется в статистической радиотехнике, то для аппроксимации (5.14) имеем $\tau_{\text{к}} = \tau_{\text{фл}}$, $\rho(\tau_{\text{к}}) \approx 0{,}4$, а для аппроксимации (5.15) $\tau_{\text{к}} = \tau_{\text{фл}} \sqrt{\pi/2} \approx 1{,}25\tau_{\text{фл}}$, $\rho(\tau_{\text{к}}) \approx 0{,}45$.

Время корреляции $\tau_{\text{к}}$ можно рассматривать как средний период замираний. Тогда средняя частота замираний — величина, обратная времени корреляции.

Многочисленные экспериментальные исследования показывают, что скорость замираний зависит от ряда факторов: частоты несущей сигнала, протяженности линии связи, диаграмм направленности антенн, типа канала и т. п. Как правило, в линиях связи большой протяженности скорость замираний значительно выше, чем в линиях с меньшей протяженностью. Средняя длительность

замираний в различных каналах может колебаться от нескольких секунд до долей секунд (0,1—0,3 с). Очень ориентировочно можно указать, что для ряда каналов скорость замираний характеризуется несколькими замираниями в секунду. Экспериментальные данные о скоростях замираний в различных линиях связи можно найти в работах [7, 10].

Важно отметить, что в подавляющем большинстве практических случаев средняя длительность замираний значительно превышает длительность отдельных посылок используемых сигналов. Это дает основание считать, что для применяемых на практике сигналов замирания являются медленными.

5.2.4. Закон распределения флюктуаций огибающей сигнала

Как уже указывалось, в каналах со случайными параметрами сигнал на входе приемника образуется в результате сложения сигналов, пришедших по разным лучам. Для удобства дальнейшего рассмотрения запишем принимаемый сигнал в несколько ином виде по сравнению с записью (5.4):

$$s_x(t)=\sum_k s_{xk}(t)=\sum_k S_{xk}(t)\sin\{\omega[t-\tau_{зk}(t)]+\varphi\}, \qquad (5.17)$$

где $s_{xk}(t)$ — сигнал, пришедший по k-му лучу; $\tau_{зk}$ — его время запаздывания; $S_{xk}(t)$ — огибающая сигнала по k-му лучу; ω и φ — частота и фаза переданного сигнала. Для упрощения записи в выражении (5.17) не учитывается явление интерференции и не отражено, какой символ был передан (x_1 или x_2), так как нас интересуют в данном случае лишь флюктуации амплитуды и фазы посылки, которые не зависят от вида переданного символа *.

Приняв во внимание (5.5), выражение (5.17) можно записать в виде

$$s_x(t)=\sum_k S_{xk}(t)\sin[\omega t+\varphi_{xk}(t)+\psi],\ 0\leqslant t\leqslant\tau_0, \qquad (5.18)$$

где $\varphi_{xk}(t)=-\omega\Delta\tau_{зk}(t)$, $\psi=\varphi-\omega\tau_{з0}$.

* В общем случае такое утверждение не верно. Однако для двоичных сигналов с активной паузой это допущение приемлемо при выполнении условия (5.11), когда замирания можно считать общими (неселективными). Далее рассматриваются только такие замирания.

На интервале, равном длительности посылки сигнала, микроструктура среды не успевает существенно измениться, так как за это время могут произойти лишь относительно небольшие случайные перемещения неоднородностей. Поэтому амплитуду сигнала по каждому лучу в течение длительности посылки практически можно считать неизменной случайной величиной, значение которой медленно изменяется от посылки к посылке. Фаза же этого сигнала даже при незначительных перемещениях неоднородностей может меняться значительно, поворачиваясь на угол 2π каждый раз, когда путь, по которому проходит этот сигнал, изменяется всего лишь на длину волны несущего колебания. Так как неоднородности на пути каждого луча перемещаются случайно, то изменения фаз сигналов, приходящих по каждому лучу, происходят также случайно.

Таким образом, сигнал, приходящий по каждому лучу, представляет собой посылку со случайной амплитудой, меняющейся медленно от посылки к посылке, и случайной фазой, которая меняется от посылки к посылке значительно быстрее, чем амплитуда. Основываясь на центральной предельной теореме и используя результаты статистической радиотехники, можно показать, что если число неоднородностей на пути луча велико, а микроструктура среды статистически однородна, то сигнал по каждому лучу можно рассматривать как стационарный случайный процесс, принадлежащий классу гауссовских случайных процессов [7, 11]. Статистические свойства огибающей такого процесса описываются законом распределения Райса (обобщенным законом Релея). В этом случае многолучевой сигнал на входе приемника (независимо от числа лучей) также будет гауссовским случайным процессом, так как композиция любого числа гауссовских процессов является также гауссовским процессом.

Если даже условия формирования гауссовского процесса по отдельному лучу не выполняются (неоднородность среды слабо выражена или рассматриваются не очень протяженные линии связи), то при суммировании большого числа сигналов отдельных лучей сигнал на входе приемника опять будет принадлежать классу гауссовских процессов. Достаточно хорошее приближение к этому получается уже при суммировании шести составляющих, что почти всегда имеет место на практике.

Представим выражение (5.18) в виде

$$s_x(t) = S_x(t)\sin[\omega t + \varphi_x(t)], \quad 0 \leqslant t \leqslant \tau_0, \qquad (5.19)$$

где постоянное значение сдвига фазы для простоты принято равным нулю; S_x и φ_x — огибающая и фаза флюктуирующего сигнала на входе приемника.

Так как случайные значения огибающей и фазы за время длительности посылки можно считать постоянными, то

$$s_x(t) = S_x \sin(\omega t + \varphi_x), \quad 0 \leqslant t \leqslant \tau_0. \qquad (5.20)$$

В некоторых случаях замирающий сигнал $s_x(t)$ может содержать регулярную составляющую $s_{\text{р}}(t)$, обусловленную тем, что помимо локальных неоднородностей на процесс распространения сигналов существенное влияние может оказать отражение и рассеяние сигналов от слоистых неоднородностей и границ раздела сред с различающимися физическими свойствами. Кроме того, при определенных условиях по некоторым направлениям возможно и когерентное распространение сигнала, когда фазы по отдельным лучам незначительно отклоняются от своего среднего значения ($|\varphi_i - \bar{\varphi}| \ll \pi$). Такие условия могут иметь место при сравнительно небольших дальностях связи или в средах со слабо выраженными неоднородностями, влияние которых не приводит к существенному случайному «расшатыванию» параметров переданного сигнала *.

С учетом сказанного представим (5.20) в виде

$$s_x(t) = S_x f_x(t) = s_{\text{р}}(t) + s_{\text{фл}}(t), \quad 0 \leqslant t \leqslant \tau_0, \qquad (5.21)$$

где $s_{\text{р}}(t)$ и $s_{\text{фл}}(t)$ — регулярная и флюктуирующая составляющие принятой посылки сигнала; $f_x(t)$ — сигнальная функция, учитывающая высокочастотное заполнение посылки и соответствующие случайные сдвиги фазы при распространении.

* Необходимо заметить, что термин «регулярная составляющая» условен, поскольку факторы, приводящие к появлению такой составляющей, нельзя считать неизменными во времени. Однако случайные изменения условий образования «регулярной составляющей» происходят сравнительно медленно и за время сеанса связи с этими изменениями часто можно не считаться. Регулярную составляющую в литературе иногда называют также «зеркальной», подчеркивая этим одну из наиболее частых причин ее появления (отражение от границы раздела сред).

Выражение (5.21) можно записать также в виде

$$s_x(t) = S_р f_р(t) + S_{фл} f_{фл}(t),\ 0 \leqslant t \leqslant \tau_0, \tag{5.22}$$

где $S_р$ и $S_{фл}$ — огибающие составляющих; $f_р(t)$ и $f_{фл}(t)$ — сигнальные функции составляющих.

Нетрудно видеть, что мощности регулярной и флюктуирующей составляющих отдельной посылки сигнала определяются выражениями

$$P_р = \frac{1}{\tau_0} \int_0^{\tau_0} s^2{}_р(t)\, dt = \frac{S^2{}_р}{\tau_0} \int_0^{\tau_0} f^2{}_р(t)\, dt = \frac{S^2{}_р}{2}, \tag{5.23}$$

$$P_{фл} = \frac{1}{\tau_0} \int_0^{\tau_0} s^2{}_{фл}(t)\, dt = \frac{S^2{}_{фл}}{\tau_0} \int_0^{\tau^0} f^2{}_{фл}(t)\, dt = \frac{S^2{}_{фл}}{2}. \tag{5.24}$$

Очевидно, что величина мощности $P_{фл}$ случайна и различна для разных посылок. Величину же $P_р$ для всех посылок можно считать одинаковой (см. сделанное выше примечание).

Мощность $P_{фл}$, усредненную по всем посылкам, назовем средней мощностью флюктуирующей составляющей сигнала и будем в дальнейшем обозначать $\bar{P}_{фл} = \bar{S}^2{}_{фл}/2$.

В соответствии с изложенными ранее соображениями мгновенное значение сигнала $s_x(t)$ на входе приемника подчиняется нормальному закону, а его огибающая S_x — обобщенному закону Релея (закону Райса), который в рассматриваемом случае можно записать в виде (см. (4.7))

$$W(S_x) = \frac{S_x}{\sigma^2{}_{фл}} \exp\left(-\frac{S^2{}_р + S^2{}_x}{2\sigma^2{}_{фл}}\right) I_0\left(\frac{S_р S_x}{\sigma^2{}_{фл}}\right),\ S_x > 0, \tag{5.25}$$

где $\sigma^2{}_{фл}$ — дисперсия случайной величины S_x.

Введем обозначение

$$\beta^2 = S^2{}_р / 2\sigma^2{}_{фл}.$$

Тогда выражение (5.25) примет вид

$$W(S_x) = \frac{S_x}{\sigma^2{}_{фл}} \exp(-\beta^2) \exp\left(-\frac{S_x}{2\sigma^2{}_{фл}}\right) \times \\ \times I_0\left(\sqrt{2}\beta \frac{S_x}{\sigma_{фл}}\right). \tag{5.26}$$

Нетрудно видеть, что величина β^2 характеризует соотношение между мощностью регулярной составляющей и средней мощностью флюктуирующей составляющей

принимаемого сигнала и с этой точки зрения может рассматриваться как регулировочный параметр распределения Райса. Действительно, учитывая (5.23) и принимая во внимание, что дисперсия $\sigma^2_{фл}$ пропорциональна средней мощности флюктуирующей составляющей $\bar{P}_{фл}$, имеем

$$\beta^2 \equiv P_р/\bar{P}_{фл}. \tag{5.27}$$

Отсюда следует, что в зависимости от соотношения между мощностями $P_р$ и $\bar{P}_{фл}$ величина β принимает значения в интервале $0 \leqslant \beta \leqslant \infty$.

Можно показать, что при больших значениях β, соответствующих средам со слабо проявляющими себя неоднородностями, закон распределения (5.26) близок к нормальному, стремясь по мере увеличения β к дельта-функции. Если величина β уменьшается, глубина флюктуаций огибающей посылок сигнала возрастает, а закон распределения этих флюктуаций приближается к закону Релея, т. е.

$$\lim_{\beta \to 0} W(S_x) = \frac{S_x}{\sigma^2_{фл}} \exp\left(-\frac{S^2_x}{2\sigma^2_{фл}}\right). \tag{5.28}$$

Это выражение следует из (5.26), если учесть, что

$$\lim_{\beta \to 0} I_0(\beta) = 1, \quad \lim_{\beta \to 0} \exp(-\beta^2) = 1.$$

В классе случайных гауссовских процессов описание замираний огибающей сигнала законом распределения Релея соответствует предельному случаю наиболее глубоких флюктуаций.

Учитывая, что средний квадрат случайной величины S_x при законе Релея ($S_р=0$) равен

$$\overline{S^2_x} = \overline{S^2_{фл}} = \int_0^\infty S^2_x W(S_x)\, dS_x = 2\sigma^2_{фл} \tag{5.29}$$

и приняв во внимание соотношение (5.24), из которого следует

$$\overline{P_{фл}} = 0{,}5\overline{S^2_{фл}} = \sigma^2_{фл},$$

выражение (5.28) можно записать в виде

$$W(S_x) = \frac{S_x}{\overline{P_{фл}}} \exp\left(-\frac{S^2_x}{2\overline{P_{фл}}}\right). \tag{5.30}$$

Здесь $\overline{P_{фл}}$ — средняя мощность флюктуирующего сигнала.

Закон Релея имеет место тогда когда ни одна составляющая многолучевого сигнала не содержит регулярной части. Канал, характеризующийся такими условиями распространения, называют релеевским. Если флюктуирующий сигнал имеет заметную регулярную часть, канал называют райсовским или квазирелеевским.

Типичными примерами релеевского канала являются многие каналы связи с рассеянной волной (тропосферные, ионосферные, метеорные и др.). К райсовским каналам можно отнести ряд каналов с прямой волной и остронаправленными диаграммами антенн (особенно приемной).

Необходимо отметить, что существенные изменения условий распространения сигналов могут приводить к тому, что один и тот же канал связи может в разное время быть либо релеевским, либо райсовским.

Механизм распространения сигналов в реальных каналах со случайными параметрами в ряде случаев значительно сложнее, чем это было рассмотрено выше. Поэтому наблюдаемые в реальных условиях флуктуации огибающей сигналов иногда характеризуются значительными изменениями их функций распределения.

Распределение Райса обладает некоторой гибкостью и с помощью параметра β позволяет учесть возможные изменения закона распределения от нормального до закона Релея. Однако наличие одного «регулировочного параметра» в распределении Райса не всегда обеспечивает необходимую гибкость в изменении вида функции распределения так, чтобы получить достаточно хорошую аппроксимацию наблюдаемого распределения. В 1943 г. японский ученый Накагами предложил довольно универсальную функцию распределения, получившую в литературе название «m-распределение» или «распределение Накагами». Эта функция имеет вид

$$W(S_x) = \frac{2}{\Gamma(m)}\left(\frac{m}{\overline{S^2_x}}\right)^m S_x^{2m-1}\exp\left(-\frac{mS^2_x}{\overline{S^2_x}}\right),\ S_x \geqslant 0,\ m \geqslant 0{,}5, \quad (5.31)$$

где $\Gamma(m)$ — гамма-функция.

Распределение Накагами оказалось весьма удобным при решении ряда задач, так как оно обладает значительной гибкостью, позволяя изменением регулировоч-

ного параметра m получать ряд важных распределений. В частности, при $m=0{,}5$ выражение переходит в нормальный закон с односторонней плотностью распределения. При $m=1$ имеем закон распределения Релея. При $m>1$ распределение Накагами становится достаточно хорошей аппроксимацией закона Райса.

Распределение Накагами позволяет описать более широкий класс распределений, чем, например, распределение Райса. По этой причине оно получило широкое распространение в статистической радиотехнике. Интересующихся этими вопросами отсылаем к работам [6, 12].

Рассмотренная упрощенная модель замираний сигнала в ряде случаев довольно хорошо согласуется с многочисленными результатами экспериментальных исследований, а также с данными практических наблюдений. Поэтому такая модель может быть положена в основу теоретических исследований как первое приближение реальных процессов в каналах со случайными параметрами.

Дальнейшее рассмотрение будем проводить в предположении, что замирания являются общими. Как уже указывалось, общие замирания не изменяют форму посылок сигнала и поэтому их можно трактовать как результат умножения сигнала на некоторую случайную функцию времени, отображающую процесс замираний в канале. Эту функцию часто называют мультипликативной случайной помехой, а соответствующие ей замирания — мультипликативными. Будем полагать, что эти замирания принадлежат к классу замираний, описываемых обобщенным законом Релея.

5.3. ОДИНОЧНЫЙ ОПТИМАЛЬНЫЙ ПРИЕМ ДВОИЧНЫХ ФЛЮКТУИРУЮЩИХ СИГНАЛОВ

Прием сигналов в каналах со случайными параметрами — значительно более сложная задача, чем прием в каналах с постоянными параметрами. Это связано с тем, что помимо неизбежного действия случайных аддитивных помех при приеме необходимо учитывать случайные изменения параметров сигнала.

Выясним, какие особенности имеет эта задача при одиночном оптимальном приеме двоичных сигналов с общими замираниями.

В § 5.2 было установлено, что характерной особенностью общих замираний является то, что амплитуда и фаза принимаемой по-

сылки сигнала — случайные величины, сохраняющие неизменное значение в пределах длительности посылки:

$$\left.\begin{matrix} S_x = \text{const} \\ 0 \leqslant S_x < \infty \end{matrix}\right\} 0 \leqslant t \leqslant \tau_0, \qquad (5.32)$$

$$\left.\begin{matrix} \varphi_x = \text{const} \\ 0 \leqslant \varphi_x \leqslant 2\pi \end{matrix}\right\} 0 \leqslant t \leqslant \tau_0. \qquad (5.33)$$

Остальные исходные данные и допущения считаем такими же, как и в гл. 3: аддитивная помеха представляет собой белый гауссовский шум; приемное устройство неискажающее; синхронизация идеальная.

Для решения задачи одиночного приема флюктуирующих сигналов можно применить теорию проверки гипотез, изложенную в гл. 3.

Пусть на вход приемника различения флюктуирующих бинарных сигналов поступает колебание

$$y(t) = s_x(t) + n(t), \quad 0 \leqslant t \leqslant \tau_0, \qquad (5.34)$$

представляющее собой аддитивную смесь флюктуирующего сигнала и нормального белого шума. В данном случае сигнал помимо переданного информационного символа x_i (полезная информация) содержит также информацию о случайных значениях амплитуды S_x и фазы φ_x (мешающая или паразитная информация). Иными словами, теперь сигнал на входе приемника представляет собой функцию времени, содержащую три неизвестных параметра: x_i, S_x и φ_x, среди которых x_i — полезный и подлежит извлечению, а S_x и φ_x — вредные, действие которых необходимо как можно сильнее ослабить.

Для того чтобы запись флюктуирующего сигнала была более наглядной и содержала в себе указанные параметры, этот сигнал удобно представить в одном из видов

$$s_i(t) = s_i(S_x, \varphi_x, t) \text{ или } s_i(t) = s(x_i, S_x, \varphi_x, t). \qquad (5.35)$$

Передача того или иного информационного символа x_i в этой записи отображается либо индексом i при s, либо самим символом x_i *.

В соответствии с теорией проверки гипотез для решения задачи оптимального приема двоичных сигналов необходимо знать апостериорные вероятности правильного приема переданных символов. Определение этих вероятностей в каналах со случайными параметрами является более трудной задачей, чем в каналах с постоянными параметрами. Если в каналах с постоянными параметрами можно было не делать никакого различия между записью апостериорной вероятности в виде $p(s_i|y)$ или $p(x_i|y)$, так как это были эквивалентные выражения, то для каналов со случайными параметрами эти выражения различны, поскольку сигнал s_i содержит еще и случайные паразитные параметры S_x и φ_x. Действие этих параметров приводит к тому, что при передаче могут появляться ошибки даже без учета влияния аддитивного шума. Поэтому теперь для принятия решения о той или иной гипотезе недостаточно вычислить только апостериорные вероятности вида $p(s_i|y)$, так как они не дают еще основания для суждения о том, какой же символ был

---

* Заметим, хотя и очевидно, что индекс x при s и φ указывает только на то, что эти параметры сигнала случайны. Никакого отношения к информационному символу x_i этот индекс не имеет.

передан. Для принятия решения необходимо определить апостериорные вероятности вида $p(x_i|y)$. Изложим методику вычисления этих вероятностей.

На основании формулы Байеса (§ 3 2) плотность распределения апостериорных вероятностей приема флюктуирующего сигнала s_i может быть записана в виде

$$W(x_i,\ S_x,\ \varphi_x|y)= \\ =W(x_i,\ S_x,\ \varphi_x)W(y|x_i,\ S_x\varphi_x)/W(y), \tag{5.36}$$

где $W(x_i,\ S_x,\ \varphi_x)$ — совместная плотность априорного распределения параметров x_i, S_x и φ_x; $W(y)$ — многомерная функция распределения входного колебания $y(t)$, представляющего собой смесь флюктуирующего сигнала и шума.

Функция $W(x_i,\ S_x,\ \varphi_x|y)$ характеризует плотность распределения вероятностей того, что данному входному колебанию $y(t)$ соответствует сигнал с параметрами x_i, S_x и φ_x. Нас же интересует вероятность того, что при данном входном $y(t)$ будет принят символ x_i, т. е. апостериорная вероятность $p(x_i|y)$. Определение этой вероятности можно выполнить в три этапа. Сначала нужно найти апостериорную плотность распределения $W(x_i,\ S_x,\ \varphi_x|y)$, затем исключить из этого распределения паразитные параметры S_x и φ_x, т. е. определить усредненное по этим параметрам распределение апостериорных вероятностей

$$W_{\overline{s\varphi}}(x_i\,|\,y)=\int\limits_{G_s G_\varphi} W(x_i,\ S_x,\ \varphi_x\,|\,y)dS_x d\varphi_x, \tag{5.37}$$

и, наконец, найти искомую апостериорную вероятность в соответствии с выражением

$$P_{\overline{s\varphi}}(x_i\,|\,y)=\int\limits_{G_{x_i}} W_{\overline{s\varphi}}(x_i\,|\,y)dx. \tag{5.38}$$

В этих выражениях G_s, G_φ — области существования случайных значений S_x и φ_x (см. (5.32) и (5.33)); G_{x_i} — область значений величины x_i; индекс $\overline{s\varphi}$ показывает, что величина получена с учетом усреднения по мешающим параметрам S_x и φ_x.

С учетом (5.36) и (5.37) выражение для апостериорной вероятности можно записать в виде

$$p_{\overline{s\varphi}}(x_i\,|\,y)= \\ =k_y\int\limits_{G_s}\int\limits_{G_\varphi}\int\limits_{G_{x_i}} W(x_i,\ S_x,\ \varphi_x)W(y\,|\,x_i,\ S_x,\ \varphi_x)\,dS_x d\varphi_x dx_i, \tag{5.39}$$

где $k_y=1/W(y)$ — величина, не зависящая от переменных интегрирования (некоторый коэффициент)

Соотношение (5.39) является исходным при анализе одиночного оптимального приема флюктуирующих сигналов. Чтобы в соответствии с этим соотношением вычислить апостериорную вероятность, нужно знать распределение $W(x_i,\ S_x,\ \varphi_x)$ и функцию правдоподобия $W(y|x_i,\ S_x,\ \varphi_x)$.

Так как случайные параметры x_i, S_x и φ_x независимы, можно записать

$$W(x_i,\ S_x,\ \varphi_x) = W(x_i)\,W(S_x)\,W(\varphi_x), \qquad (5.40)$$

где $W(x_i)$ — априорное распределение плотности вероятностей посылки символа x_i; $W(S_x)$ и $W(\varphi_x)$ — плотности распределения случайных мешающих параметров S_x и φ_x соответственно.

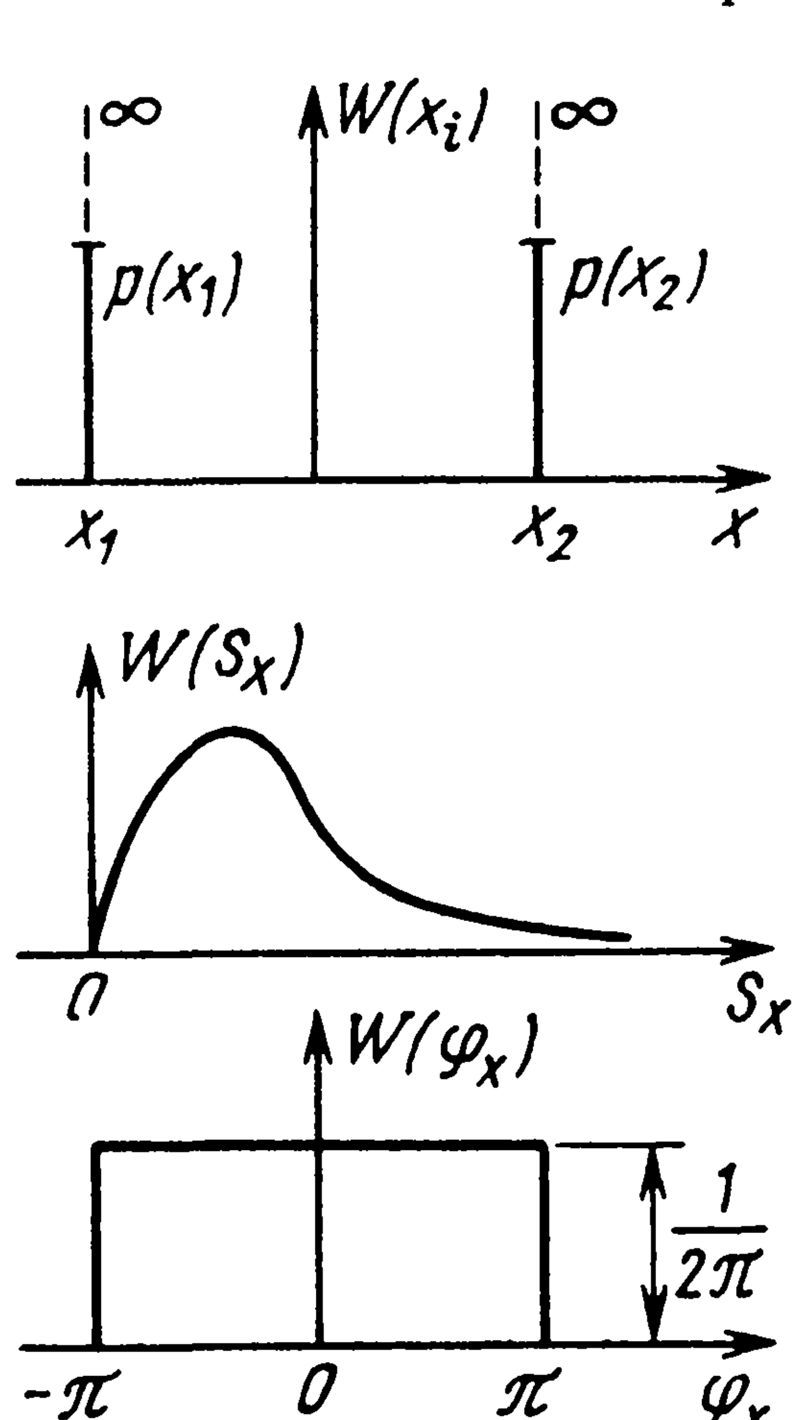


Рис. 5.6.

Полагается, что распределение (5.40) известно, так как оно представляет собой априорные сведения о статистических свойствах источника дискретных сообщений и канала связи. Вид отдельных априорных распределений для равновероятной передачи символов и релеевского канала связи показан на рис. 5.6. Распределение $W(x_i)$ имеет вид двух дельта-функций с весами $p(x_1)=p(x_2)=$ $=0{,}5$.

Для нахождения функции правдоподобия воспользуемся следующими соображениями. Аддитивная помеха $n(t)$ типа внутреннего флюктуационного шума не имеет никакой статистической связи с передаваемыми дискретными сообщениями x_i, а также с параметрами канала. Следовательно, при фиксированных значениях параметров x_i, S_x и φ_x функция правдоподобия полностью определяется многомерным распределением нормального шума (§ 3.2)

$$W(y|x_i,\ S_x,\ \varphi_x) = = W_m\,[y(t) - s(x_i,\ S_x,\ \varphi_x,\ t)] \qquad (5.41)$$

или с учетом выражения (3.22)

$$W(y\,|\,x_i,\ S_x,\ \varphi_x) = = k_1 \exp\left\{-\frac{1}{N_0}\int_0^{\tau_0} [y(t) - s(x_i,\ S_x,\ \varphi_x,\ t)]^2 dt\right\}, \qquad (5.42)$$

где k_1 — постоянный коэффициент в выражении многомерного распределения белого гауссова шума (см. (3.22)).

Располагая выражениями (5.40) и (5.42), соотношение (5.39) для искомых апостериорных вероятностей можно записать в виде

$$p_{\overline{s\varphi}}(x_i\,|\,y) = k \int_{G_s}\int_{G_\varphi}\int_{G_{x_i}} W(x_i)W(S_x)W(\varphi_x)\,\times$$

$$\times \exp\left\{-\frac{1}{N_0}\int_0^{\tau_0} [y(t) - s(x_i,\ S_x,\ \varphi_x,\ t)]^2 dt\right\} dx_i dS_x d\varphi_x, \qquad (5.43)$$

где $k=k_1k_y$ — коэффициент.

При вычислении этого выражения нужно иметь в виду, что плотность распределения дискретной случайной величины x_i, которая может принимать значения x_1 или x_2, представляет собой совокупность δ-функций с весами, равными априорным вероятностям этих величин:

$$W(x_i) = \sum_i p(x_i)\delta(x - x_i). \qquad (5.44)$$

Выражение (5.43) представляет собой запись апостериорных вероятностей в общем виде. Естественно, что если нужно найти вероятность $p_{\overline{s\varphi}}(x_1|y)$, то интегрирование по переменной x_i нужно вести только в той области ее значений, где она равна x_1. При этом

$$\int\limits_{G_{x_1}} W(x_i)dx_i = \int\limits_{G_{x_1}} p(x_i)\delta(x - x_i)dx_i = p(x_1). \qquad (5.45)$$

Для нахождения вероятности $p_{\overline{s\varphi}}(x_2 \mid y)$ интегрирование нужно вести по области значений $x_i = x_2$.

С учетом сказанного окончательно имеем следующие выражения апостериорных вероятностей:

$$p_{\overline{s\varphi}}(x_1 \mid y) = kp(x_1) \int\limits_{G_s} \int\limits_{G_\varphi} \exp\left\{-\frac{1}{N_0}\int\limits_0^{\tau_0} [y(t) - s(x_1, S_x,\right.$$

$$\left.\varphi_x, t)]^2 dt\right\} W(S_x)W(\varphi_x)dS_x d\varphi_x, \qquad (5.46)$$

$$p_{\overline{s\varphi}}(x_2 \mid y) = kp(x_2) \int\limits_{G_s} \int\limits_{G_\varphi} \exp\left\{-\frac{1}{N_0}\int\limits_0^{\tau_0} [y(t) - s(x_2, S_x,\right.$$

$$\left.\varphi_x, t)]^2 dt\right\} W(S_x)W(\varphi_x)dS_x d\varphi_x. \qquad (5.47)$$

После определения этих вероятностей дальнейший анализ одиночного оптимального приема для каналов со случайными параметрами выполняется в такой же последовательности, как и для каналов с постоянными параметрами. В соответствии с найденными апостериорными вероятностями нужно взять логарифм отношения

$$\ln \Lambda_{12} = \ln[p_{\overline{s\varphi}}(x_1 \mid y)/p_{\overline{s\varphi}}(x_2 \mid y)], \qquad (5.48)$$

а затем с учетом байесовского правила решения определить алгоритм работы оптимального приемника и установить его структуру. Анализ завершается нахождением выражения для вероятности ошибки при передаче одного информационного символа.

Необходимо отметить, что задача оптимального различения флюктуирующих сигналов из-за математических трудностей сложна для исследования и ее решение удается довести до конца только в некоторых частных случаях.

Исследования вопросов одиночного оптимального приема флюктуирующих сигналов, выполненные в работах [14—17], показывают, что структура приемников значительно усложняется по

сравнению с оптимальным приемом сигналов, известных точно, а помехоустойчивость падает.

Построение оптимальных приемников для одиночного приема флюктуирующих сигналов с инженерной точки зрения часто оказывается нецелесообразным, так как при значительном усложнении такие приемники не позволяют сколь-нибудь существенно повысить помехоустойчивость по сравнению с неоптимальным приемом.

По указанным причинам ограничимся изложенной выше методикой решения задачи одиночного оптимального приема флюктуирующих сигналов и не будем рассматривать приложения этой методики к различным конкретным способам передачи.

Детальное рассмотрение вопросов оптимального одиночного приема флюктуирующих сигналов и синтеза оптимальных приемников в соответствии с изложенной методикой можно найти в работах [12—17]. Некоторые из указанных вопросов читателю предлагается решить самостоятельно (см. задачи к этой главе).

Чтобы убедиться в невысокой эффективности одиночного оптимального приема по сравнению с неоптимальным, достаточно рассмотреть и сопоставить их помехоустойчивость.

5.4. ВЕРОЯТНОСТЬ ОШИБКИ ПРИ ОДИНОЧНОМ ПРИЕМЕ ФЛЮКТУИРУЮЩИХ СИГНАЛОВ

Выше (§ 5.2) указывалось, что общие (мультипликативные) замирания могут служить приемлемой моделью, которая учитывает основные особенности реальных процессов распространения в ряде каналов со случайными параметрами. Допущение о мультипликативности замираний позволяет сравнительно просто решить задачу определения вероятности ошибки при одиночном приеме флюктуирующих сигналов. Рассмотрим эту задачу.

Если бы параметры среды были постоянны, то отдельную посылку на входе приемника можно было бы записать в виде

$$s(t)=S_0\sin(\omega t+\varphi),\quad 0\leqslant t\leqslant\tau_0, \qquad (5.49)$$

где значения амплитуды огибающей S_0 и постоянного сдвига фазы полностью определяются геометрией распространения сигнала (дальностью, диграммами направленности и т. п.) и постоянными физико-химическими свойствами данной среды распространения (средней величиной поглощения энергии сигнала на единицу расстояния, средней скоростью распространения колебаний в среде и т. п.).

Если параметры микроструктуры среды случайны и замирания сигнала, обусловленные этими параметрами, можно считать общими, посылка сигнала на входе приемника определяется выражением (5.20), которое можно представить в виде

$$s_x(t) = \frac{S_x}{S_0} S_0 \sin(\omega t + \varphi_x), \quad 0 \leqslant t \leqslant \tau_0. \tag{5.50}$$

Обозначим

$$\mu = S_x/S_0. \tag{5.51}$$

Тогда

$$s_x(t) = \mu S_0 \sin(\omega t + \varphi_x), \quad 0 \leqslant t \leqslant \tau_0. \tag{5.52}$$

Сопоставляя выражения (5.49) и (5.52), заключаем, что для модели общих замираний влияние случайных параметров среды на переданную посылку сигнала можно учесть случайным мультипликативным коэффициентом μ и аддитивным случайным сдвигом фазы φ_x несущей этой посылки.

Величину μ иногда называют коэффициентом передачи среды со случайными параметрами [16]. Этот коэффициент, учитывающий влияние микроструктуры среды, однозначно связан линейной зависимостью со случайной величиной S_x и, следовательно, подчиняется такому же закону распределения, что и величина S_x.

Случайные значения μ и φ_x за время длительности посылки остаются неизменными, а их сколь-нибудь заметные изменения происходят на интервалах времени, значительно превышающих эту длительность. Основываясь на этом, можно определить вероятность ошибки приема любой отдельной посылки флюктуирующего сигнала, пользуясь результатами, полученными в гл. 3 и 4 для каналов с постоянными параметрами. Обозначим эту вероятность $p_{\text{ош}}(h_x)$, где h_x — величина, связанная с отношением энергии данной посылки флюктуирующего сигнала к спектральной плотности аддитивного шума. Так как

$$h^2_x = E_x/N_0 = S^2_x \tau_0/2N_0, \tag{5.53}$$

где E_x — энергия посылки флюктуирующего сигнала, то

$$h_x = S_x \sqrt{\tau_0/2N_0} = \mu S_0 \sqrt{\tau_0/2N_0}. \tag{5.54}$$

Величина h_x является случайной, поэтому вероятность $p_{\text{ош}}(h_x)$ нужно рассматривать как условную вероятность, которая соответствует данному значению h_x. Для получе-

ния безусловной вероятности ошибки одиночного приема флюктуирующего сигнала необходимо усреднить вероятность $p_{ош}(h_x)$ по всем возможным значениям случайной величины h_x, т. е. найти величину

$$\overline{p}_{ош}=\int\limits_0^\infty p_{ош}(h_x)\,W(h_x)\,dh_x. \qquad (5.55)$$

В этом выражении $W(h_x)$ — закон плотности распределения величины h_x.

Выражение (5.55) определяет среднюю вероятность ошибки приема одиночной посылки флюктуирующего сигнала

Применим изложенную методику к определению вероятности ошибки при оптимальных и неоптимальных способах приема посылок сигнала в каналах с релеевским законом замирании. Ограничимся рассмотрением таких замираний по следующим причинам:

1) замирания по закону Релея соответствуют наиболее глубоким флюктуациям огибающей сигнала; следовательно, анализ позволит получить верхнюю (наихудшую) границу помехоустойчивости флюктуирующих сигналов *;

2) для ряда каналов со случайными параметрами замирания по закону Релея довольно типичны;

3) анализ помехоустойчивости при релеевских флюктуациях наиболее прост.

Для определения вероятности ошибки в соответствии с выражением (5.55), необходимо знать закон плотности распределения случайной величины h_x. Найдем этот закон.

Представим (5.53) в виде

$$h^2{}_x=\frac{S^2{}_x\tau_0\Delta f_э}{2N_0\Delta f_э}=\frac{S^2{}_x}{2\sigma^2{}_ш}Б_с=\frac{P_{cx}}{P_ш}Б_с, \qquad (5.56)$$

где $\sigma^2{}_ш=P_ш$ — средняя мощность шума на входе приемника, учитываемая в полосе пропускания приемника $\Delta f_э$; P_{cx} — мощность посылки сигнала с амплитудой S_x; $Б_с=\Delta f_э\tau_0$ — база сигнала.

Из (5.56) следует

$$h_x=S_x\sqrt{Б_с}/\sqrt{2}\,\sigma_ш. \qquad (5.57)$$

---

* В литературе имеются указания, что иногда замирания могут быть более глубокими, чем это следует из закона Релея [16]. Эти случаи довольно редки, и мы на них останавливаться не будем.

Так как закон распределения случайной величины S_x известен, то нетрудно найти и закон распределения величины h_x. Для этого можно воспользоваться теоремой о функциональном преобразовании случайных величин и их распределений [11] В соответствии с этой теоремой плотность вероятности случайной величины y, связанной со случайной величиной x однозначной функциональной зависимостью $y=f(x)$, определяется выражением

$$W(y)=W_x[\varphi(y)]\,|\varphi'(y)|, \tag{5.58}$$

где $\varphi(y)=x$ — функция, обратная функции $f(x)$; $\varphi'(y)$ — производная функция $\varphi(y)$; $W_x(x)$ — закон распределения случайной величины x.

Приняв во внимание выражение (5.28) для распределения величины S_x и выполнив несложные преобразования в соответствии с (5.58), получим распределение величины h_x в виде

$$W(h_x)=\frac{2h_x}{\overline{h^2{}_x}}\exp\left(-\frac{h^2{}_x}{\overline{h^2{}_x}}\right), \tag{5.59}$$

где

$$\overline{h^2{}_x}=\frac{\sigma^2{}_{\text{фл}}}{\sigma^2{}_{\text{ш}}}Б_{\text{c}}=\left(\frac{\overline{P_{\text{фл}}}}{P_{\text{ш}}}\right)_{\text{вх}}Б_{\text{c}}=\frac{\overline{E_{\text{фл}}}}{N_0}\cdot \tag{5.60}$$

В свою очередь,

$$\overline{E_{\text{фл}}}=\overline{P_{\text{фл}}}\tau_0 \tag{5.61}$$

— средняя энергия посылки флюктуирующего сигнала.

Полученное распределение (5.59) является релеевским. Это естественно, поскольку случайные величины h_x и S_x связаны линейной зависимостью. Значение $\overline{h^2{}_x}/2$ можно рассматривать как дисперсию случайной величины h_x. Если учесть, что величину $h^2{}_x$, определяемую выражением (5.56), можно представить в виде

$$h^2{}_x=\frac{\mu^2 S^2{}_0}{2\sigma^2{}_{\text{ш}}}Б_{\text{c}}=\mu^2\left(\frac{P_{\text{c}}}{P_{\text{ш}}}\right)_{\text{вх}}Б_{\text{c}}=\mu^2 h^2, \tag{5.62}$$

где h^2 — величина, характеризующая энергетическое отношение сигнал/шум на входе приемника без учета случайных параметров среды, то нетрудно показать, что

$$\overline{h^2{}_x}=\overline{\mu^2}h^2. \tag{5.63}$$

Из физических соображений естественно положить $\overline{\mu^2}=1$, поскольку на входе приемника средняя энергия посылки флюктуирующего сигнала должна быть равна энергии посылки сигнала в отсутствие флюктуаций.

Тогда выражение (5.59) можно записать в форме, которая и будет использоваться нами в дальнейшем:

$$W(h_x) = \frac{2h_x}{h^2} \exp\left(-\frac{h^2_x}{h^2}\right). \qquad (5.64)$$

Перейдем к определению вероятности ошибки при одиночном приеме флюктуирующих сигналов. Рассмотрение проведем для двух случаев.

1. Оптимальный когерентный прием посылки сигнала. Условная вероятность ошибки приема отдельной посылки сигнала для разных видов манипуляции определяется выражением (3.107), которое в данном случае принимает вид

$$p_{\text{ош}}(h_x) = 0{,}5 - \Phi(\gamma_c h_x). \qquad (5.65)$$

Подставив (5.64) и (5.65) в формулу (5.55), имеем

$$\overline{p_{\text{ош}}} = \int\limits_0^\infty [0{,}5 - \Phi(\gamma_c h_x)]\, \frac{2h_x}{h^2} \exp\left(-\frac{h^2_x}{h^2}\right) dh_x. \qquad (5.66)$$

Определение вероятности ошибки сводится к вычислению этого интеграла

Приняв во внимание, что

$$\frac{2h_x}{h^2} \exp\left(-\frac{h^2_x}{h^2}\right) dh_x = \exp\left(-\frac{h^2_x}{h^2}\right) d\left(\frac{h^2_x}{h^2}\right), \qquad (5.67)$$

имеем

$$\overline{p_{\text{ош}}} = 0{,}5 - \int\limits_0^\infty \Phi(\gamma_c h_x) \exp\left(-\frac{h^2_x}{h^2}\right) d\left(\frac{h^2_x}{h^2}\right). \qquad (5.68)$$

Обозначив

$$u = \Phi(\gamma_c h_x), \quad dv = \exp(-h^2_x/h^2)\, d(h^2_x/h^2)$$

и имея в виду, что

$$du = \frac{d\Phi(\gamma_c h_x)}{dh_x} = \frac{\gamma_c}{\sqrt{2\pi}} \exp\left(-\frac{\gamma^2_c h^2_x}{2}\right),$$

нетрудно взять интеграл по частям. Опуская промежуточные выкладки, запишем окончательный результат вычисления в виде

$$\overline{p_{\text{ош}}} = 0{,}5\left(1 - \gamma_c h/\sqrt{2 + \gamma^2_c h^2}\right). \qquad (5.69)$$

Для больших значений h выражение можно упростить, если иметь в виду, что

$$\frac{1}{\sqrt{2 + \gamma^2_c h^2}} \approx \frac{1}{\gamma_c h}\left(1 - \frac{1}{\gamma^2_c h^2}\right). \qquad (5.70)$$

С учетом (5.70) имеем

$$\overline{p_{\text{ош}}} \approx 1/2\gamma^2{}_c h^2. \qquad (5.71)$$

При $\gamma_c h > 4{,}5$ погрешность приближенного выражения не превышает 5%.

Формула (5.71) характеризует помехоустойчивость оптимального приема двоичных сигналов в релеевском канале.

В гл. 3 было показано, что при условии ограниченной пиковой мощности посылок сигнала величина коэффициента γ_c равна

$$\gamma_{c\text{АМ}} = 1/\sqrt{2}, \ \gamma_{c\text{ЧМ}} = 1, \ \gamma_{c\text{ФМ}} = \sqrt{2}. \qquad (5.72)$$

Соответственно вероятность ошибки для разных видов манипуляции и оптимальном приеме определяется выражениями

$$\overline{p_{\text{ош}}} = 1/h^2 \quad \text{для АМн,} \qquad (5.73)$$

$$\overline{p_{\text{ош}}} = 1/2h^2 \quad \text{для ЧМн,} \qquad (5.74)$$

$$\overline{p_{\text{ош}}} = 1/4h^2 \quad \text{для ФМн.} \qquad (5.75)$$

Обсуждение полученных результатов проведем после рассмотрения второго случая.

2. Реальный некогерентный прием посылки сигнала. В этом случае условная вероятность ошибки приема отдельной посылки сигнала для разных видов манипуляции определяется выражением (4.80), в котором вместо h нужно рассматривать случайную величину h_x:

$$p_{\text{ош}}(h_x) = 0{,}5 \exp(-\gamma^2{}_c h^2{}_x/2). \qquad (5.76)$$

Поступая так же как и в первом случае, запишем выражение для средней вероятности ошибки в релеевском канале. С учетом (5.64) и (5.76) оно примет вид

$$p_{\text{ош}} = \int_0^\infty \frac{h_x}{h^2} \exp\left(-\frac{h^2{}_x}{h^2}\right) \exp\left(-\frac{\gamma^2{}_c h^2{}_x}{2}\right) dh_x \qquad (5.77)$$

или

$$\overline{p_{\text{ош}}} = \frac{1}{2}\int_0^\infty \exp\left[-\frac{h^2{}_x}{h^2}\left(1+\frac{\gamma^2{}_c h^2}{2}\right)\right] d\left(\frac{h^2{}_x}{h^2}\right). \qquad (5.78)$$

Интегрирование этого выражения дает

$$\overline{p_{\text{ош}}} = 1/(2+\gamma^2{}_c h^2). \qquad (5.79)$$

Для больших значений h имеем приближенную формулу

$$\overline{p_{\text{ош}}} \approx 1/\gamma^2{}_c h^2. \quad (5.80)$$

Погрешность не более 5% обеспечивается при $\gamma_c h > 6,5$. Учитывая значения коэффициента γ_c, можно записать

$$\overline{p_{\text{ош}}} = 2/h^2 \quad \text{для АМн}, \quad (5.81)$$

$$\overline{p_{\text{ош}}} = 1/h^2 \quad \text{для ЧМн}, \quad (5.82)$$

$$\overline{p_{\text{ош}}} = 1/2h^2 \quad \text{для ФМн}. \quad (5.83)$$

Из полученных результатов следует, что в релеевском канале сохраняются такие же энергетические соотношения между разными видами сигналов и способами приема, как и в канале с постоянными параметрами. Наиболее помехоустойчивы сигналы с манипуляцией фазы на π, а наименее помехоустойчивы сигналы с амплитудной манипуляцией. Для каждого вида сигнала когерентный прием обеспечивает двойной энергетический выигрыш по сравнению с некогерентным. Однако помехоустойчивость одиночного приема в релеевском канале значительно ниже, чем в гауссовском. Для получения малых вероятностей ошибки в релеевском канале необходимо создавать значительное превышение сигнала над шумом. Например, если для достижения вероятности $p_{\text{ош}} = 10^{-5}$ в гауссовском канале при некогерентном приеме необходимо обеспечивать значения h^2, равные

$$h^2{}_{\text{АМ}} = 45, \ h^2{}_{\text{ЧМ}} = 22, \ h^2{}_{\text{ФМ}} = 11,$$

то соответствующие значения h^2 в релеевском канале должны быть равны

$$h^2{}_{\text{АМ}} \approx 2 \cdot 10^5, \ h^2{}_{\text{ЧМ}} \approx 10^5, \ h^2{}_{\text{ФМ}} \approx 5 \cdot 10^4.$$

По указанной причине одиночный прием двоичных сигналов в релеевских каналах не находит применения. Следовательно, не представляет практического интереса и отыскание структур оптимальных приемников для одиночного приема сигналов в таких каналах.

Одним из возможных путей существенного повышения качественных показателей передачи цифровой информации в релеевских каналах является применение разнесенного приема сигналов. Рассмотрению основных особенностей метода разнесенного приема и способов его реализации посвящены следующие параграфы этой главы.

5.5. РАЗНЕСЕНИЕ СИГНАЛОВ КАК МЕТОД БОРЬБЫ С ЗАМИРАНИЯМИ. ВИДЫ РАЗНЕСЕНИЯ

Разнесение сигналов является эффективным методом повышения качества передачи информации в каналах со случайными параметрами. Суть метода состоит в том, что решение о приеме сообщения принимается на основе анализа нескольких различающихся образцов сигнала, несущих одно и то же сообщение *. При передаче цифровой информации по каналам со случайными параметрами эти образцы представляют собой флюктуирующие копии переданной посылки сигнала.

Допустим, что имеется n различных флюктуирующих некоррелированных образцов (копий) посылки сигнала, содержащих одну и ту же информацию, т. е. соответствующих одному и тому же элементу передаваемой цифровой последовательности. Полагая замирания общими, обозначим образцы, соответствующие одной переданной посылке сигнала, через

$$s_{x1}(t) = S_{x1} f_{x1}(t), \ldots; s_{xj}(t) = S_{xj} f_{xj}(t), \ldots; s_{xn}(t) = S_{xn} f_{xn}(t), \quad (5.84)$$

где S_{xj} — огибающая образца с номером j, а $f_{xj}(t)$ — сигнальная функция этого образца. Все эти образцы действуют на интервале времени $0 \leqslant t \leqslant \tau_0$.

Пусть вероятность того, что амплитуда флюктуирующего образца с номером j меньше некоторой величины U_0, равна p_j: $p(S_{xj} < U_0) = p_j$. Тогда вероятность того, что все n образцов будут иметь амплитуды меньше величины U_0, равна

$$p(S_{xi} < U_0, S_{x2} < U_0, \ldots, S_{xn} < U_0) = \prod_{j=1}^{n} p_j. \quad (5.85)$$

Если статистика замираний всех образцов одинакова, то $p_1 = p_2 = \ldots = p_j = \ldots = p_n = p$ и вероятность сложного события (5.85) равна

$$p_{\text{общ}} = \prod_{j=1}^{n} p_j = p^n. \quad (5.86)$$

* Метод разнесения сигналов является одним из примеров принципа дублирования, который находит широкое применение в различных областях техники для повышения надежности функционирования разнообразных устройств, механизмов и систем.

Из этого выражения следует, что с увеличением числа образцов n вероятность одновременных сильных замираний всех образцов резко уменьшается. Например, $p_{общ} = 10^{-3}$ при $p_1 = 0{,}1$ и $n = 3$. Отмеченное обстоятельство можно использовать для ослабления влияния замираний при передаче цифровой информации в каналах со случайными параметрами.

Чтобы реализовать эту возможность, необходимо располагать нужным числом образцов принимаемого сигнала.

В настоящее время известны следующие виды разнесения, позволяющие получить флюктуирующие копии (образцы) одного и того же сигнала: частотное, временное, угловое, поляризационное и пространственное. Рассмотрим кратко суть этих видов разнесения сигналов.

Частотное разнесение сигналов. При этом виде необходимые для приема образцы сигнала создаются одновременной передачей одной и той же информации на разных частотах, разнесенных так, чтобы обеспечивалась слабая коррелированность образцов между собой. По сути дела, при частотном разнесении сигналов одна и та же информация одновременно передается по нескольким параллельным каналам. Нетрудно видеть, что такое разнесение сигналов требует значительного расширения используемого частотного диапазона и связано с усложнением как приемной, так и передающей аппаратуры системы связи. Поэтому частотное разнесение не нашло широкого применения при передаче информации в каналах со случайными параметрами *.

Временное разнесение сигналов. Суть этого вида разнесения состоит в том, что передача одной и той же информации повторяется через некоторые интервалы времени, соизмеримые или превышающие время корреляции замираний $\tau_к$. Естественно, что необходимость повторения передачи одной и той же информации приводит к уменьшению скорости ее передачи.

Поскольку при общих замираниях должно выполняться условие $\tau_к \gg \tau_0$, при реализации временного разнесения нужно использовать устройства памяти достаточно большого объема как в передающей, так и в приемной частях системы и значительно усложнять аппаратуру.

---

* В некоторых случаях, например для СВЧ линий связи прямой видимости [5] или для связи между подвижными объектами [16], частотное разнесение может оказаться целесообразным.

Снижение скорости передачи и усложнение аппаратуры систем при временном разнесении сигналов являются основными причинами того, что этот вид разнесения не получил широкого применения на практике.

Угловое разнесение сигналов. Этот вид разнесения основан на применении приемных антенн с остронаправленными диаграммами, ориентированными в направлении ожидаемого прихода сигналов от передатчика. При этом между отдельными диаграммами создается некоторый разнос по углу, обеспечивающий получение различающихся образцов сигнала.

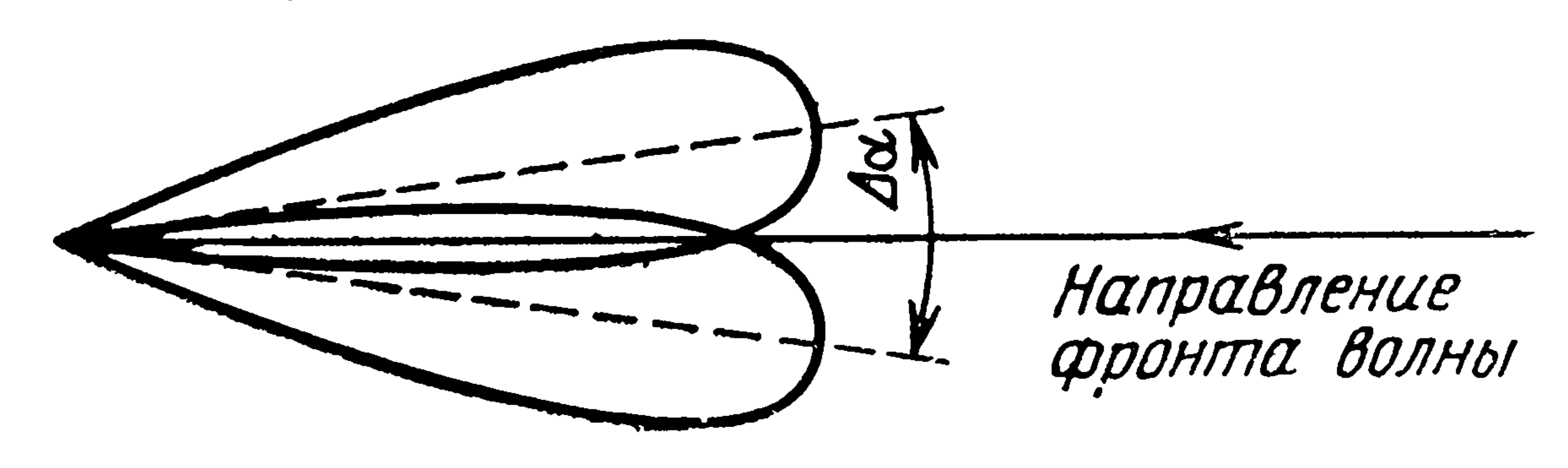


Рис. 5.7.

Угловое разнесение иллюстрируется рис. 5.7. Из рассмотрения этого рисунка следует, что при увеличении угла разноса Δα корреляция между образцами будет уменьшаться, а их уровень падать. Если же угол разноса Δα уменьшать, то керреляция между образцами будет возрастать и образцы станут мало различающимися между собой. Следовательно, при угловом разнесении должны существовать оптимальные углы, соответствующие наиболее эффективному разнесению. Исследования показывают [5, 10], что угловое разнесение в системах ионосферной и тропосферной связи может быть довольно эффективным средством борьбы с замираниями Этот вид разнесения может применяться для связи между подвижными объектами, когда нельзя применить другие виды разнесения, или как дополнительная мера при использовании комбинированных способов ослабления влияния замираний сигналов.

Поляризационное разнесение сигналов. В ряде случаев поляризация излученной волны при прохождении среды со случайными параметрами может изменяться и в точке приема волна имеет составляющие с вертикальной и горизонтальной поляризацией. Исследования показывают, что прием этих составляющих на две антенны с различной поляризацией позволяет получить образцы со сравнительно слабой корреляцией. Необходимо отме-

тить, что заметный поляризационный эффект наблюдается только в некоторых каналах [5, 10, 18]. Поэтому такой вид разнесения имеет ограниченное применение и не может рассматриваться как эффективное средство борьбы с замираниями.

Пространственное разнесение сигналов. Этот вид разнесения основан на том, что один и тот же сигнал одновременно принимается на несколько антенн, расположенных в пространстве на некотором расстоянии друг от друга. При определенном расстоянии разноса антенн корреляция между огибающими полученных копий сигнала может быть ослаблена до необходимой величины.

Итак, для ослабления влияния замираний необходимо создавать несколько флюктуирующих копий одного и того же сигнала, полученных разнесением сигналов одним из указанных способов так, чтобы обеспечить малую корреляцию между огибающими этих копий.

Многочисленные исследования показали, что коэффициент корреляции между огибающими копий сигнала зависит от величины разнесения и может быть аппроксимирован выражением [16, 19]

$$\rho(\Delta x)=\exp\left(-\Delta x^2/2\Delta x^2_0\right), \qquad (5.87)$$

где Δx — величина интервала разноса (по частоте, времени, углу, расстоянию); Δx_0 — характеристический интервал разноса, определяемый для каждого вида разнесения и конкретных условий работы экспериментально. При разносе, равном характеристическому интервалу ($\Delta x=\Delta x_0$),

$$\rho(\Delta x_0)=1/\sqrt{e}\simeq 0{,}6. \qquad (5.88)$$

Исследования показали [5, 10], что увеличение интервала разноса сверх характеристического не приводит к существенному улучшению качества приема флюктуирующих сигналов. Поэтому применять слишком большие интервалы разноса практически нецелесообразно.

Анализ, выполненный в работах [5, 7, 20], показывает, что при достаточно больших отношениях сигнал/шум на входе приемника ($\overline{h^2}_x=h^2\gg 1$) влияние корреляции между образцами сигналов эквивалентно некоторому уменьшению средней мощности образца. Это уменьшение можно оценить коэффициентом

$$\gamma_\rho=1/\sqrt{1-\rho}. \qquad (5.89)$$

В работе [20] показано, что если корреляция между любой парой образцов одинакова, то коэффициент, характеризующий уменьшение средней мощности образца, определяется выражением

$$\gamma_\rho \approx 1/(1-\rho^2). \qquad (5.90)$$

Для значений $\rho<0,6$ выражения (5.89) и (5.90) дают приблизительно одинаковые результаты. Поэтому при оценках можно пользоваться любым из них.

Если $\rho\leqslant 0,4$—0,5, то энергетический проигрыш из-за наличия корреляции между образцами незначителен (менее 1 дБ). Отсюда следует важный практический вывод, что разносить сигналы для уменьшения коэффициента корреляции ниже величины 0,4—0,5 нецелесообразно.

Необходимо отметить, что при угловом и пространственном разнесении сигналов корреляция между сигналами зависит не только от абсолютной величины разноса, но и от геометрии разнесения. Так, при пространственном разнесении сигналов, образованных рассеянной волной (тропосферное, ионосферное и т. п. распространение) коэффициент корреляции с увеличением расстояния разноса между антеннами, расположенными в вертикальном или горизонтальном направлении, перпендикулярном трассе, обычно падает быстрее, чем вдоль трассы.

Для УКВ линий с прямой волной наилучшим является вертикальное разнесение, разнесение же в других направлениях практически невыполнимо, так как для него требуются чрезмерно большие расстояния разноса [5]. При благоприятном направлении пространственного разноса антенн соотношение между длиной волны λ сигнала и расстоянием разноса Δl, обеспечивающим значение коэффициента корреляции $\rho\leqslant 0,4$—0,5, приблизительно определяется величиной $\Delta l/\lambda \sim 10$—20. Отсюда следует, что достаточная декорреляция флюктуирующих образцов достигается при разнесении антенн на расстояния порядка одного-двух десятков длин волн.

Пространственное разнесение широко применяется в каналах со случайными параметрами. Поэтому дальнейшее рассмотрение будем вести применительно к этому виду разнесения.

5.6. ОСНОВНЫЕ МЕТОДЫ РАЗНЕСЕННОГО ПРИЕМА

Одной из главных задач приема разнесенных сигналов является такая обработка полученных копий флюктуирующего сигнала, которая обеспечивает извлечение переданного сообщения с наименьшими «затратами», понимаемыми в том или ином смысле. В одних случаях наиболее важным фактором оказывается простота технической реализации разнесенного приема, в других — необходимая величина мощности излучения передатчика для обеспечения заданного качества приема информации и т. д. Поэтому указанная задача не имеет однозначного решения и одинаковые показатели качества принятой информации могут быть достигнуты различными методами.

Прежде чем рассматривать эти методы и их особенности, сформулируем задачу разнесенного приема в общей постановке, а также введем ряд допущений и ограничений, при которых будет проведено дальнейшее исследование.

Пусть с помощью пространственного разнесения получены n флюктуирующих копий одного и того же сигнала, каждая из которых поступает на вход свой ветви разнесения. Под ветвью разнесения понимают совокупность устройств, предназначенных для выполнения тех или иных операций с отдельной копией сигнала. В каждой ветви имеются внутренние флюктуационные шумы, которые накладываются на принятые копии сигналов. С учетом влияния этих шумов колебания на входах ветвей разнесения можно представить в виде

$$\begin{gathered} y_1(t)=s_{x1}(t)+n_1(t), \\ \dots\dots\dots\dots\dots \\ y_j(t)=s_{xj}(t)+n_j(t), \\ \dots\dots\dots\dots\dots \\ y_n(t)=s_{xn}(t)+n(t), \end{gathered} \tag{5.91}$$

где $j=1,\ 2,\dots,n$ — номер ветви разнесения; $s_{xj}(t)$ — копия сигнала на входе j-й ветви разнесения; $n_j(t)$ — внутренние шумы j-й ветви разнесения, пересчитанные к ее входу.

Для выделения переданного сообщения выходы всех ветвей подключаются к схеме объединения, которая работает по некоторому заранее установленному правилу.

Результат объединения поступает на вход устройства анализа и принятия решения о том, какой информационный символ был передан. Сказанное поясняется структурной схемой, приведенной на рис. 5.8.

Известные в настоящее время методы разнесенного приема можно разделить на две группы: 1) додетекторное объединение ветвей: 2) последетекторное объединение ветвей. Рассмотрим особенности указанных групп разнесенного приема.

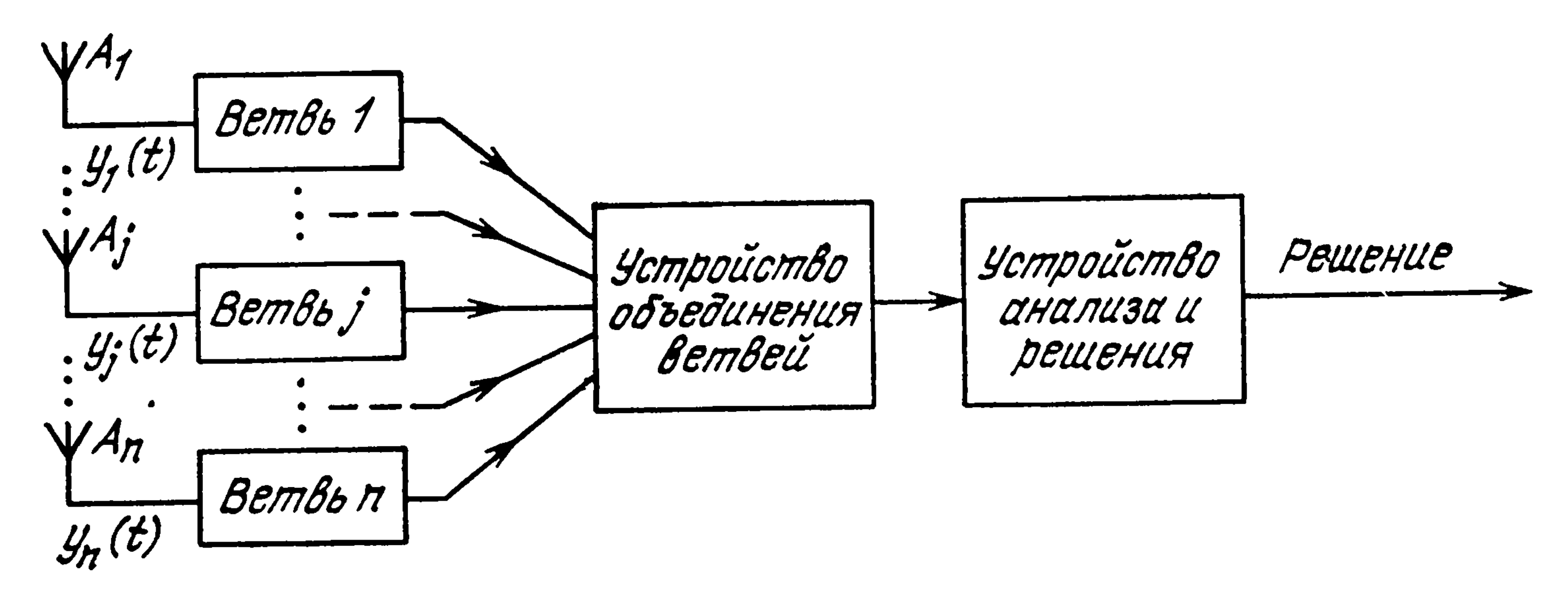


Рис. 5.8.

5.6.1. *Додетекторное объединение ветвей*

Особенностью этой группы методов является объединение ветвей на частоте заполнения копии с последующим анализом полученного колебания. При этом обработка копий в отдельных ветвях разнесения сводится только к усилению и фильтрации принятых копий от помех в полосе частот, занимаемой спектром сигнала.

В результате объединения отдельных «зашумленных» образцов $y_j(t)$ образуется колебание $y_р(t)$, которое будем называть результирующим колебанием. Это колебание представляет собой некоторую композицию «зашумленных» флюктуирующих копий сигнала. Полученное колебание можно рассматривать как сумму двух колебаний

$$y_р(t) = s_{xр}(t) + n_р(t), \tag{5.92}$$

где $s_{xр}(t)$ — сигнальная часть результирующего колебания; $n_р(t)$ — его шумовая часть, обусловленная влиянием внутренних шумов отдельных ветвей разнесения.

Колебание $y_р(t)$ далее поступает на устройство анализа, которое по сути дела выполняет роль приемного устройства для такого колебания и должно решать задачу различения сигнала $s_{xр}(t)$ на фоне шума $n_р(t)$. Так как при передаче двоичной цифровой информации сигнал $s_{xр}(t)$ может содержать информацию либо о символе x_1, либо о символе x_2, то указанное устройство в результате анализа колебания $y_р(t)$ должно принять решение о том, какой символ принят и выдать это решение получателю.

Выясним некоторые принципиальные особенности разнесенного приема с додетекторным объединением ветвей.

Как и ранее, будем считать замирания сигнала, а следовательно, и его копий, общими. Обозначим отношение мощности сигнала к мощности шума в отдельной ветви через

$$h^2_{xj}=P_{сxj}/P_{шj}=S^2_{xj}/2\sigma^2_{шj}, \qquad (5.93)$$

где $P_{шj}=\sigma^2_{шj}$ — средний квадрат шума (мощность шума) в j-й ветви, определяемый выражением

$$P_{шj}=\sigma^2_{шj}=\frac{1}{\tau_0}\int\limits_0^{\tau_0} n^2_j(t)\,dt=\overline{n^2_j(t)}, \qquad (5.94)$$

а $P_{сxj}$ — мощность посылки сигнала в j-й ветви, определяемая выражением

$$P_{сxj}=\frac{1}{\tau_0}\int\limits_0^{\tau_0} s^2_{xj}(t)\,dt=\frac{1}{\tau_0}\int\limits_0^{\tau_0} S^2_{xj}f^2_{xj}(t)\,dt=\frac{S^2_{xj}}{2}. \qquad (5.95)$$

Величина h^2_{xj} является случайной, поскольку величина огибающей S_{xj} каждой копии случайна. Однако если усреднить эту величину с учетом закона распределения $W(S_{xj})$ огибающей S_{xj}, полученное значение окажется уже не случайным и будет характеризовать указанное отношение в среднем. Обозначим его через

$$\overline{h^2_{xj}}=h^2_j=\overline{S^2_{xj}}/2\sigma^2_{шj}=\overline{P_{сxj}}/\sigma^2_{шj}, \qquad (5.96)$$

где

$$\overline{P_{сxj}}=0{,}5\int\limits_0^{\infty} S^2_{xj}\,W(S_{xj})\,dS_{xj}=\overline{S^2_{xj}}/2. \qquad (5.97)$$

— средняя мощность флюктуирующей копии сигнала.

Введем аналогичные обозначения для результирующего колебания $y_р(t)$ на выходе схемы объединения ветвей

$$h^2{}_{xр}=P_{cxр}/P_{шр}=S^2{}_{xр}/2\sigma^2{}_{шр} \quad (5.98)$$

и

$$\overline{h^2{}_{xр}}=h^2{}_р=\overline{S^2{}_{xр}}/2\sigma^2{}_{шр}, \quad (5.99)$$

где

$$\overline{S^2{}_{xр}}=\int_0^\infty S^2{}_{xр}W(S_{xр})\,dS_{xр}. \quad (5.100)$$

Здесь $W(S_{xр})$ — закон распределения огибающей сигнальной части результирующего колебания.

Величина $h^2{}_{xр}$, так же, как и величина $h^2{}_{xj}$, является случайной. Очевидно, что эта величина и закон ее распределения зависят от числа ветвей, способа (правила) их объединения, статистики замирания копий сигнала и уровня шума в отдельных ветвях. Очевидно также, что применение разнесенного приема эффективно, если только оно обеспечивает энергетический выигрыш в отношении сигнал/шум после объединения ветвей и «улучшает» закон распределения замирания огибающей результирующего колебания, делая эти замирания менее глубокими по сравнению с замираниями в отдельных ветвях.

Таким образом, необходимо применять такие способы додетекторного объединения копий разнесенного сигнала, которые дают эффект по двум направлениям:

1. Улучшают отношение средней мощности сигнала к мощности шума на выходе схемы объединения по сравнению с соответствующим отношением в отдельных ветвях, т. е. обеспечивают условие

$$\overline{h^2{}_{xр}}=h^2{}_р>\overline{h^2{}_{xj}}=h^2{}_j. \quad (5.101)$$

Это условие можно заменить эквивалентным условием вида

$$B_р=\overline{h^2{}_{xр}}/\overline{h^2{}_{xj}}=h^2{}_р/h^2{}_j>1. \quad (5.102)$$

В теории разнесенного приема величину $B_р$ часто называют энергетическим выигрышем разнесения.

2. Изменяют в благоприятную сторону закон распределения огибающей результирующего сигнала по срав-

нению с законом распределения огибающих в отдельных ветвях, т. е. обеспечивают условие

$$p(S_{xp} \leqslant U_0) \ll p(S_{xj} \leqslant U_0) \qquad (5.103)$$

или эквивалентное ему условие

$$\int_0^{U_0} W(S_{xp})\, dS_{xp} \ll \int_0^{U_0} W(S_{xj})\, dS_{xj}, \qquad (5.104)$$

где U_0 — некоторый пороговый уровень; $p(S_{xp} \leqslant U_0)$ и $p(S_{xj} \leqslant U_0)$ — вероятности того, что величины S_{xp} и S_{xj} окажутся ниже порога U_0.

Из проведенного рассмотрения следует, что задача разнесенного приема с додетекторным объединением ветвей (с объединением на частоте заполнения посылок сигнала) сводится к решению трех вопросов:

1) нахождению эффективных способов объединения ветвей, обеспечивающих указанные выше улучшения;

2) определению способов анализа колебания, полученного после объединения ветвей, с целью извлечения из него переданной информации;

3) определению вероятности ошибки при разнесенном приеме флюктуирующего сигнала.

Очевидно, что возможен ряд вариантов разнесенного приема, которые могут отличаться методами объединения ветвей и способами анализа результирующего колебания $y_p(t)$. При оптимальном методе объединения ветвей и оптимальном способе анализа колебания $y_p(t)$ разнесенный прием обеспечивает минимальную вероятность ошибки. Такой прием называется оптимальным разнесенным приемом при додетекторном объединении ветвей.

В настоящее время наиболее часто применяют линейные методы объединения ветвей, которые характеризуются тем, что результирующее колебание на выходе схемы объединения представляет собой линейную комбинацию выходных колебаний отдельных ветвей:

$$y_p(t) = \sum_{j=1}^{n} k_j y_j(t) = \sum_{j=1}^{n} k_j s_{xj}(t) + \sum_{j=1}^{n} k_j n_j(t) \qquad (5.105)$$

$$y_p(t) = s_{xp}(t) + n_p(t), \qquad (5.106)$$

где

$$s_{x\text{р}}(t)=\sum_{j=1}^{n}k_j s_{xj}(t)=\sum_{j=1}^{n}k_j S_{xj} f_{xj}(t) \qquad (5.107)$$

представляет сигнальную часть колебания $y_\text{р}(t)$ на выходе схемы объединения, а

$$n_\text{р}(t)=\sum_{j=1}^{n}k_j n_j(t) \qquad (5.108)$$

— его шумовую часть. В этих выражениях k_j — некоторый весовой коэффициент, величина которого зависит от конкретного метода объединения ветвей.

Из линейных методов объединения ветвей наибольшее распространение в теории и технике разнесенного приема получили следующие: а) автовыбор наилучшей ветви; б) простое линейное сложение ветвей; в) оптимальное линейное сложение ветвей.

При разнесенном приеме каждый из этих методов может сочетаться с одним из возможных способов анализа результирующего колебания $y_\text{р}(t)$. Как указывалось выше, анализ этого колебания эквивалентен задаче одиночного приема флюктуирующего сигнала $s_{x\text{р}}(t)$ на фоне шума $n_\text{р}(t)$. Такая задача может быть решена различными способами приема. При оптимальном приеме задачу можно решить, применяя методику, изложенную в § 5.3. Это позволит определить структуру оптимального устройства анализа колебания $y_\text{р}(t)$. При реальном приеме можно применить один из способов некогерентного приема, рассмотренных в гл. 4.

Вероятность ошибки приема сигнала $s_{x\text{р}}(t)$ характеризует помехоустойчивость разнесенного приема. Для определения этой вероятности необходимо знать закон распределения огибающей сигнала $S_{x\text{р}}$ на выходе схемы объединения ветвей. Зная этот закон и учитывая соотношение (5.58), можно найти закон распределения случайной величины $h_{x\text{р}}$. Поскольку замирания считаются общими, вероятность ошибки в случае разнесенного приема можно найти, применяя методику, рассмотренную в § 5.4. В соответствии с этой методикой искомая вероятность определяется выражением

$$\overline{p_{\text{ош}\text{р}}}=\int_0^\infty p_\text{ош}(h_{x\text{р}})\,W(h_{x\text{р}})\,dh_{x\text{р}}. \qquad (5.109)$$

Анализ разнесенного приема, основанного на линейных методах объединения ветвей, изложен в § 5.7.

5.6.2. Последетекторное объединение ветвей

Известен ряд методов разнесенного приема, в которых используется последетекторное объединение ветвей. Особенностью этой группы методов является то, что в каждой ветви производится не только усиление и фильтрация принятой копии сигнала, но и анализ этой копии. Этот анализ может быть выполнен либо полностью, включая принятие решения о том, какому информационному символу соответствует данная копия, либо частично — без принятия такого решения. Окончательное решение о принятом символе выносится после объединения ветвей и анализа результатов, полученных в отдельных ветвях.

В соответствии с изложенным правилом работы разнесенного приема с последетекторным объединением ветвей возможен ряд вариантов такого приема. Эти варианты различаются в зависимости от того, как выполнен анализ в каждой ветви (оптимально или неоптимально) и доведен этот анализ до принятия решения или решение в ветвях не принято.

В настоящее время в литературе рассмотрен ряд методов разнесенного приема с последетекторным объединением ветвей [16—20]. Наиболее известны из них метод квадратичного сложения ветвей и метод дискретного сложения ветвей. Суть этих методов и их особенности рассмотрены в § 5.8.

В дальнейшем при рассмотрении различных методов разнесенного приема будем полагать, что все ветви разнесения имеют одинаковые параметры и одинаковые средние мощности сигнала и шума (т. е. однородны):

$$\overline{P_{сx1}} = \overline{P_{сx2}} = \ldots = \overline{P_{сxj}} = \ldots = \overline{P_{сxn}} = P_с, \qquad (5.110)$$

$$\overline{n^2_1(t)} = \ldots = \overline{n^2_j(t)} = \ldots = \overline{n^2_n(t)} = \Delta f_э N_0 = P_ш, \qquad (5.111)$$

где $\Delta f_э$ — полоса пропускания ветви, определяемая шириной спектра посылки сигнала; N_0 — спектральная плотность флюктуационного шума.

5.7. МЕТОДЫ РАЗНЕСЕННОГО ПРИЕМА С ДОДЕТЕКТОРНЫМ ЛИНЕЙНЫМ ОБЪЕДИНЕНИЕМ ВЕТВЕЙ

При использовании методов разнесенного приема с линейным объединением ветвей предполагается, что копии переданной посылки сигнала одновременно поступают на входы отдельных ветвей. Это означает, что время многолучевого растяжения τ_p отдельных копий при распространении сигнала незначительно по сравнению с длительностью посылки τ_0 и при анализе вопросов разнесенного приема эффектом многолучевого растяжения можно пренебречь. Ограничения, к которым приводит такое допущение, будут рассмотрены после анализа методов разнесенного приема.

5.7.1. Разнесенный прием с использованием автовыбора лучшей ветви

Особенностью этого метода приема является то, что устройство объединения анализирует отношение сигнал/шум в каждой ветви и выбирает ветвь, в которой это отношение максимально. При одинаковых параметрах ветвей (усиление, полоса пропускания и уровень шума) лучшей в этом смысле всегда будет ветвь, в которой максимальна огибающая сигнала. Так как огибающая сигнала в каждой ветви случайна, то спустя некоторое время ее величина в выбранной ветви может стать меньше некоторого порогового уровня, а в одной из других ветвей может заметно возрасти и стать больше этого уровня. Тогда устройство объединения должно отключить предыдущую выбранную ветвь и подключить вновь выбранную.

Итак, при автовыборе из всех n случайных величин $h_{x1}, h_{x2}, \ldots, h_{xj}, \ldots, h_{xn}$ или эквивалентных им величин $S_{x1}, S_{x2}, \ldots, S_{xj}, \ldots, S_{xn}$, выбирается максимальная на данном интервале анализа величина h_x или S_x. Такой выбор означает, что в выражении (5.105) все весовые коэффициенты k_j принимаются равными нулю, за исключением одного, который берется равным единице.

$$k_j = \begin{cases} 1 \text{ для ветви с } S_{xj} = \max S_x, \\ 0 \text{ для других ветвей.} \end{cases}$$

В соответствии с этим огибающая сигнала на выходе схемы объединения представляет собой случайную величину, равную

$$S_{xp} = \max S_{xj},$$

где номер j может меняться случайно от одного интервала анализа состояния ветвей к другому.

Сказанное поясняется рис. 5.9, где показаны огибающие двух флюктуирующих копий сигнала и огибающая сигнала S_{xp} на выходе схемы объединения. Для наглядности эта огибающая показана более толстой линией.

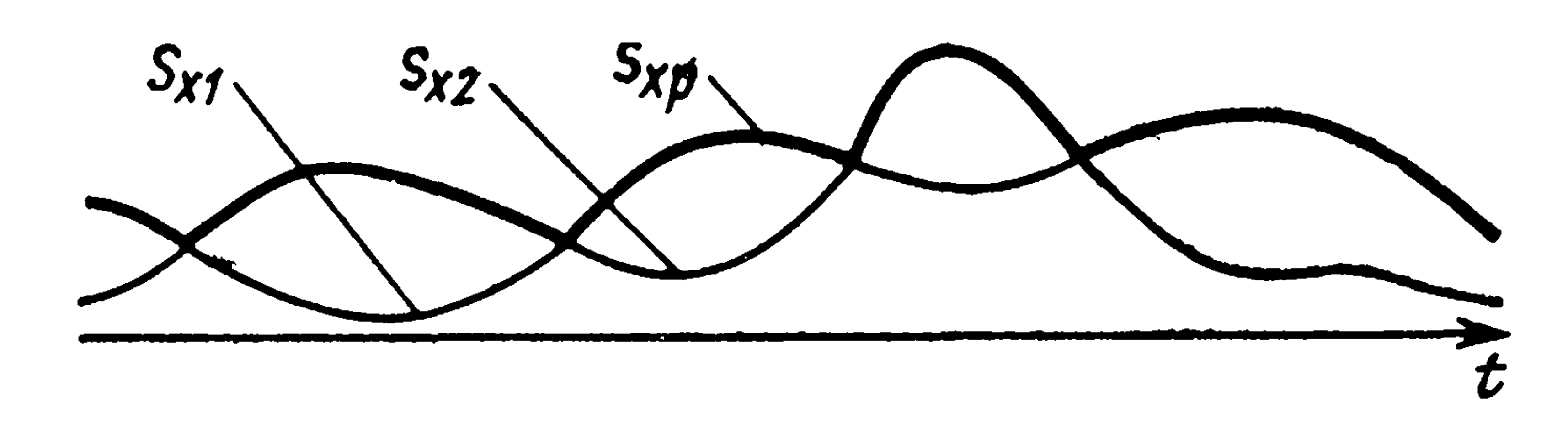


Рис. 5.9.

Поскольку при автовыборе в создании выходного колебания всегда участвует только одна из n ветвей, мощность шумовой составляющей этого колебания равна (см. (5.108) и (5.111))

$$P_{шр} = \overline{n^2_p(t)} = \sigma^2_{шр} = \sigma^2_ш = N_0 \Delta f_э. \qquad (5.112)$$

Найдем закон распределения огибающей S_{xp}, считая, что все ветви разнесения (копии сигнала) независимы. Запишем вероятность того, что в некоторый момент времени значение огибающей сигнала в j-ветви окажется меньше значения огибающей результирующего сигнала, в виде

$$p(S_{xj} < S_{xp}) = \int_0^{S_{xp}} W_1(S_{xj})\, dS_{xj} = F_1(S_{xp}), \qquad (5.113)$$

где $W_1(S_{xj})$ — одномерный закон плотности распределения случайной величины S_{xj}; $F_1(S_{xp})$ — интегральная функция распределения случайной величины S_{xj}.

Вероятность сложного события, состоящего в том, что ни в одной из ветвей огибающая сигнала не превысит величину S_{xp} и может стать только равной ей, определяется (при независимых и одинаковых ветвях) выражением

$$p = \prod_{j=1}^{n} p(S_{xj} < S_{xp}) = [F_1(S_{xp})]^n. \qquad (5.114)$$

Это выражение представляет собой интегральную функцию распределения сложного события S_{xp}

$$F(S_{xp}) = [F_1(S_{xp})]^n. \tag{5.115}$$

Взяв производную выражения (5.115) по S_{xp} и приняв во внимание, что

$$\frac{dF_1(S_{xp})}{dS_{xp}} = W_1(S_{xp}),$$

получим

$$\frac{dF(S_{xp})}{dS_{xp}} = W(S_{xp}) = nW_1(S_{xp})[F_1(S_{xp})]^{n-1}. \tag{5.116}$$

Выражение (5.116) определяет плотность распределения вероятностей огибающей сигнала на выходе устройства объединения при автовыборе. Оно справедливо для любого закона распределения замираний огибающих сигнала в отдельных ветвях (при условии, что они одинаковы и независимы).

Если замирания в отдельных ветвях определяются законом Релея (см. (5.28)), то

$$W_1(S_{xp}) = \frac{S_{xp}}{\sigma^2_{фл}} \exp\left(-\frac{S^2_{xp}}{2\sigma^2_{фл}}\right), \tag{5.117}$$

$$F_1(S_{xp}) = 1 - \exp(-S^2_{xp}/2\sigma^2_{фл}), \tag{5.118}$$

где $\sigma^2_{фл}$ — средняя мощность флюктуирующего сигнала в отдельной ветви (см. (5.28)).

С учетом (5.117) и (5.118) выражение (5.116) принимает вид

$$W(S_{xp}) = n\frac{S_{xp}}{\sigma^2_{фл}} \exp\left(-\frac{S^2_{xp}}{2\sigma^2_{фл}}\right)\left[1 - \exp\left(-\frac{S^2_{xp}}{2\sigma^2_{фл}}\right)\right]^{n-1}. \tag{5.119}$$

Приняв во внимание (5.119) и (5.98), а также методику, рассмотренную в § 5.4, нетрудно от распределения величины S_{xp} перейти к распределению величины h_{xp}

$$W(h_{xp}) = n\frac{2h_{xp}}{h^2} \exp\left(-\frac{h^2_{xp}}{h^2}\right)\left[1 - \exp\left(-\frac{h^2_{xp}}{h^2}\right)\right]^{n-1}, \tag{5.120}$$

где $h^2 = \overline{h^2_{xj}}$ — отношение средней мощности сигнала к средней мощности шума в отдельной ветви.

Отношение средних мощностей сигнала к шуму на выходе схемы объединения можно найти по формуле

$$\overline{h^2{}_{xp}} = \int\limits_0^\infty h^2{}_{xp} W(h_{xp})\, dh_{xp},$$

которая с учетом (5.120) принимает вид

$$h^2{}_{xp} = nh^2 \int\limits_0^\infty \frac{h^2{}_{xp}}{h^2} \exp\left(-\frac{h^2{}_{xp}}{h^2}\right)\left[1 - \right.$$

$$\left. - \exp\left(-\frac{h^2{}_{xp}}{h^2}\right)\right]^{n-1} d\left(\frac{h^2{}_{xp}}{h^2}\right). \quad (5.121)$$

Использовав разложение в бином Ньютона, интеграл можно взять по частям. В результате имеем

$$\overline{h^2{}_{xp}} = h^2 \sum_{k=1}^{n} 1/k. \quad (5.122)$$

Из полученного выражения следует, что энергетический выигрыш при автовыборе определяется формулой

$$B_{a} = \overline{h^2{}_{xp}}/h^2 = \sum_{k=1}^{n} 1/k. \quad (5.123)$$

Из этой формулы видно, что величина выигрыша связана с числом ветвей разнесения нелинейной зависимостью: увеличение числа ветвей не приводит к пропорциональному увеличению выигрыша.

Найдем теперь вероятность ошибки при использовании рассматриваемого способа приема. Будем считать, что последующая обработка колебания, полученного на выходе схемы объединения, осуществляется одним из некогерентных способов. Приняв во внимание выражения (5.76), (5.120) и (5.109), имеем

$$\overline{p_{\text{ош}}} = n \int\limits_0^\infty \frac{h_{xp}}{h^2} \exp\left[-\frac{h^2{}_{xp}}{h^2}\left(1 + \frac{h^2\gamma^2{}_{c}}{2}\right)\right] \times$$

$$\times \left[1 - \exp\left(-\frac{h^2{}_{xp}}{h^2}\right)\right]^{n-1} dh_{xp}. \quad (5.124)$$

Как и в предыдущем случае, этот интеграл можно взять по частям, предварительно разложив одну из подынтегральных функций в бином Ньютона. После взятия интеграла и некоторых преобразований получаем следу-

ющее выражение для вероятности ошибки при автовыборе

$$\overline{p_{\text{ош}}} = \frac{n!}{2} \prod_{k=1}^{n} \frac{2}{(2k + \gamma^2_{\text{c}} h^2)}. \qquad (5.125)$$

При $n=1$ формула (5.125) полностью совпадает с выражением (5.79), полученным для одиночного некогерентного приема.

Если среднее значение отношения сигнал/шум в ветвях достаточно велико ($h^2 \gg 1$), формулу (5.125) можно упростить:

$$\overline{p_{\text{ош}}} \approx \frac{n!}{2} \left(\frac{2}{\gamma^2_{\text{c}} h^2} \right)^n. \qquad (5.126)$$

Из этого выражения следует, что применение разнесенного приема позволяет существенно уменьшить вероятность ошибки по сравнению с одиночным приемом флюктуирующих сигналов. Так, например, при $\gamma_{\text{c}} h = 10$ вероятность ошибки одиночного приема $p_{\text{ош}} = 10^{-2}$, а разнесенного (при $n=2$) $\overline{p_{\text{ош}}} = 4 \cdot 10^{-4}$, т. е. в 25 раз меньше.

Известны два варианта реализации разнесенного приема с использованием автовыбора ветвей: с переключением приемников отдельных ветвей и с переключением антенн.

Рассмотренный выше анализ относится к первому варианту. При переключении антенн схема оказывается проще, так как достаточно иметь один приемник, который подключается к той или иной антенне. Анализ этого варианта проводить не будем. Исследования показывают, что он дает несколько худшие результаты, чем вариант с переключением приемников [21].

При реализации того или иного варианта автовыбора необходимо иметь в виду неизбежные переходные процессы, возникающие при переключении приемников или антенн. Стремление ослабить влияние этих процессов на достоверность принимаемой информации приводит к необходимости применять более широкую по сравнению с оптимальной полосу пропускания, что естественно снижает отношение сигнал/шум в отдельных ветвях и ухудшает предельные возможности разнесенного приема с автовыбором. К некоторым особенностям приема с автовыбором мы еще вернемся в конце этого параграфа при сравнении различных способов линейного объединения ветвей.

5.7.2. Разнесенный прием с линейным сложением ветвей

Характерной особенностью этого метода приема является то, что при образовании результирующего колебания все ветви считаются равноценными ($k_j=1$ для всех ветвей), а объединение копий сигнала происходит когерентно. Для того чтобы обеспечить когерентное (линейное) объединение ветвей, необходимо все копии сигнала в отдельных ветвях сфазировать, т. е. добиться условий, при которых фазы этих копий станут одинаковыми. Такие условия можно создать, применив систему автоподстройки фаз ($n-1$) копий к фазе одной из них. Учитывая (5.105) и (5.106), для линейного объединения ветвей с одинаковыми параметрами имеем

$$y_p(t)=\sum_{j=1}^{n} S_{xj} f_j(t)+\sum_{j=1}^{n} n_j(t).] \tag{5.127}$$

В результате фазирования сигнальные функции во всех ветвях становятся одинаковыми, т. е. $f_1(t)=$ $=f_2(t)= \ldots =f_j(t)= \ldots =f_n(t)$. Обозначив их через $f_s(t)$, выражение (5.127) можно записать в виде

$$y_p(t)=f_s(t)\sum_{j=1}^{n} S_{xj}+\sum_{j=1}^{n} n_j(t) \tag{5.128}$$

или, приняв во внимание (5.51), в виде

$$y_p(t)=S_0 f_s(t)\sum_{j=1}^{n} \mu_j+\sum_{j=1}^{n} n_j(t), \tag{5.129}$$

где μ_j — коэффициент передачи среды для j-й ветви.

Сигнальная часть результирующего колебания $y_p(t)$ определяется выражением

$$s_{xp}(t)=f_s(t)\sum_{j=1}^{n} S_{xj}=S_0 f_s(t)\sum_{j=1}^{n} \mu_j, \tag{5.130}$$

в котором огибающая равна

$$S_{xp}=\sum_{j=1}^{n} S_{xj}=S_0\sum_{j=1}^{n} \mu_j. \tag{5.131}$$

Из выражения (5.131) следует, что нахождение закона распределения огибающей $S_{\text{хр}}$ сводится к композиции законов распределения огибающих S_{xj} в отдельных ветвях или к композиции законов распределения коэффициентов μ_j. Если замирания сигнала в отдельных ветвях определяются законом Релея, то задача сводится к композиции n распределений Релея. К сожалению, решение этой задачи связано с весьма громоздкими выкладками и его не удается выразить через известные функции. Даже для $n=2$ распределение имеет довольно сложный вид [11, 16]. Мы этими вопросами заниматься не будем. Позже мы убедимся, что особой необходимости в нахождении указанного закона распределения нет. Поэтому ограничимся лишь определением величины энергетического выигрыша при линейном объединении ветвей.

Нетрудно видеть, что в соответствии с (5.98) для рассматриваемого способа отношение мощности сигнала к мощности шума на выходе схемы объединения определяется случайной величиной

$$h^2_{\text{хр}} = \frac{S^2_0}{2n\sigma^2_{\text{ш}}}\left(\sum_{j=1}^{n}\mu_j\right)^2 = \frac{S^2_0}{2n\sigma^2_{\text{ш}}}\left[\sum_{j=1}^{n}\mu^2_j + \sum_{\substack{i=1 \\ (j\neq i)}}^{n}\sum_{j=1}^{n}\mu_i\mu_j\right]. \quad (5.132)$$

Среднее значение этой величины равно

$$\overline{h^2_{\text{хр}}} = \frac{h^2}{n}\left[\overline{\sum_{j=1}^{n}\mu^2_j} + \overline{\sum_{j=1}^{n}\sum_{j=1}^{n}\mu_i\mu_j}\right],\ j\neq i,$$

где

$$h^2 = \overline{h^2_x} = S^2_0/2\sigma^2_{\text{ш}} = \overline{P_{\text{c}xj}}/P_{\text{ш}j}$$

— отношение средней мощности сигнала к мощности шума в отдельной ветви. Так как копии сигнала в отдельных ветвях считаются независимыми, то среднее значение произведения независимых величин равно произведению их средних значений и, следовательно,

$$\overline{h^2_{\text{хр}}} = \frac{h^2}{n}\left[\overline{\sum_{j=1}^{n}\mu^2_j} + \overline{\sum_{j=1}^{n}\mu_j}\,\overline{\sum_{i=1}^{n}\mu_i}\right],\ i\neq j. \quad (5.133)$$

Если замирания сигнала в отдельных ветвях определяются законом Релея, то, учитывая выражения (5.51) и (5.58), нетрудно показать, что распределение случай-

ной величины μ_j также подчиняется закону Релея и имеет вид

$$W(\mu_j)=2\mu_j \exp(-\mu^2_j). \qquad (5.134)$$

С учетом этого выражения

$$\overline{\left(\sum_{j=1}^{n} \mu^2_j\right)}=\sum_{j=1}^{n} \int_0^\infty 2\mu^3_j \exp(-\mu^2_j)\, d\mu_j = n,$$

$$\overline{\left(\sum_{i=1}^{n} \mu_i\right)}=\overline{\left(\sum_{j=1}^{n} \mu_j\right)}=\sum_{j=1}^{n} \int_0^\infty 2\mu^2_j \exp(-\mu^2_j)\, d\mu_j = n\sqrt{\pi}/2.$$

В результате выражение (5.133) принимает вид

$$\overline{h^2_{xp}}=\frac{h^2}{n}\left[n+n(n-1)\frac{\pi}{4}\right]=h^2\left[1+(n-1)\frac{\pi}{4}\right]. \qquad (5.135)$$

Отсюда следует, что величина энергетического выигрыша при линейном объединении ветвей определяется формулой

$$B_л=\overline{h^2_{xp}}/h^2=1+(n-1)\pi/4\approx 0{,}22+0{,}78n, \qquad (5.136)$$

из которой видно, что выигрыш пропорционален числу ветвей.

Этот результат показывает, что с энергетической точки зрения линейное объединение ветвей более эффективно, чем автовыбор лучшей ветви.

Исследования показывают, что для $n=2$ вероятность ошибки при разнесенном приеме с линейным сложением ветвей $\overline{p_{ош\,л}}$ незначительно отличается от вероятности ошибки при разнесенном приеме с автовыбором $\overline{p_{ош\,а}}$ [21]. Для оценки вероятности ошибки при $n>2$ можно пользоваться приближенным соотношением

$$\overline{p_{ош\,л}}\approx B_а p_{ош\,а}/B_л,$$

где $B_а$ и $B_л$ — величина выигрыша при автовыборе и линейном сложении соответственно.

С точки зрения технической реализации разнесенный прием с линейным сложением ветвей сравнительно несложен. На рис. 5.10 приведена упрощенная структурная схема приемной части с двумя ветвями разнесения. В этой схеме автоматическая регулировка усиления (АРУ) обеспечивает одинаковое усиление в ветвях. Когерентность (линейность) сложения ветвей достигается с помощью системы ФАП, приводящей фазу одного сиг-

нала к фазе другого. После объединения ветвей обработка полученного колебания $y_р(t)$ ничем не отличается от одиночного приема и в зависимости от применяемого вида передачи (вида манипуляции сигнала) может быть выполнена одним из способов приема, рассмотренных в гл. 4.

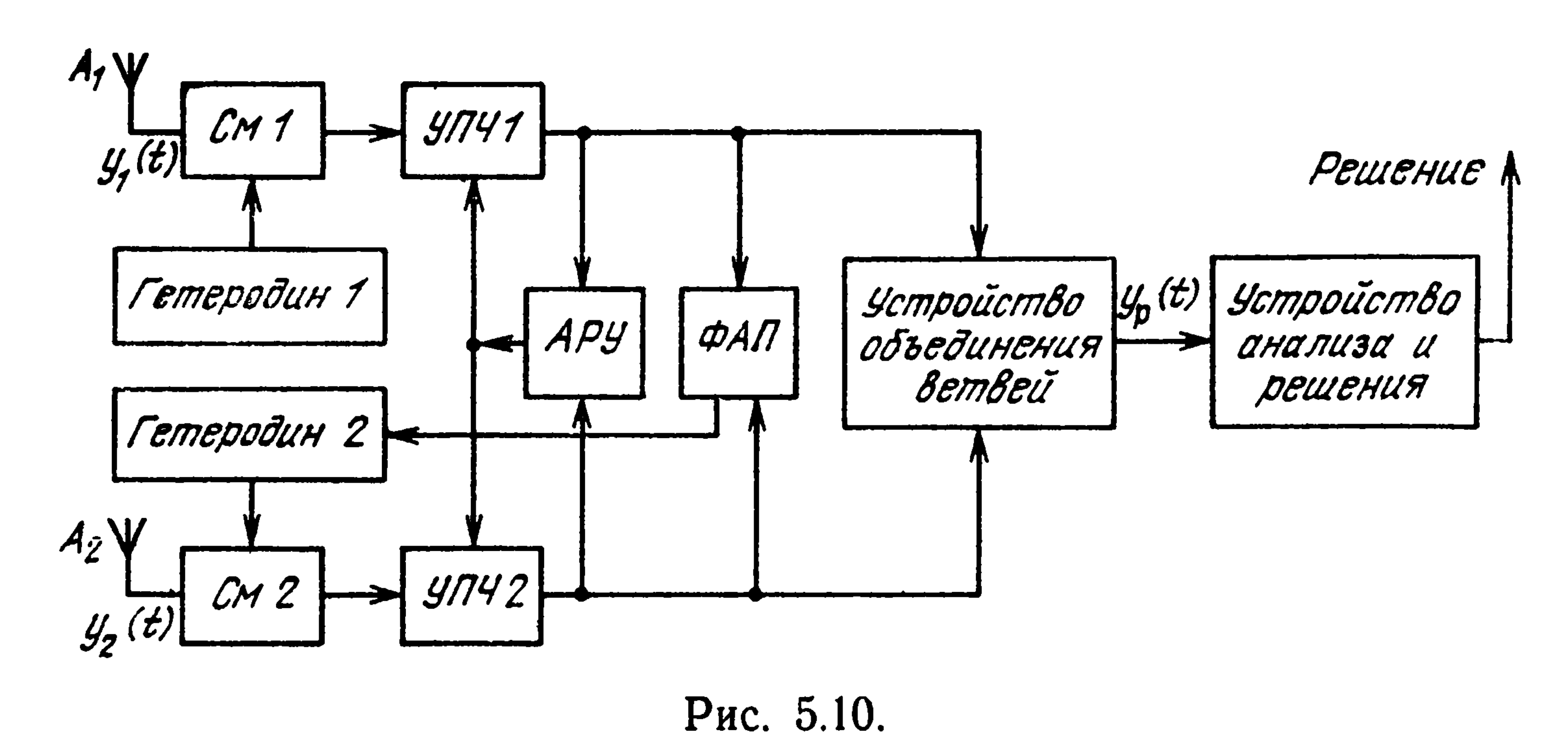


Рис. 5.10.

5.7.3. Разнесенный прием с оптимальным линейным сложением ветвей

При линейном сложении все ветви считаются равноценными, что учитывается одинаковыми весовыми коэффициентами независимо от фактического отношения сигнал/шум в каждой ветви Принципиальным недостатком такого метода объединения является то, что ветви с плохим отношением сигнал/шум вносят заметный вклад в шумовую составляющую результирующего колебания и незначительный — в сигнальную составляющую. Очевидно, что с этой точки зрения линейное объединение ветвей с одинаковыми весовыми коэффициентами неоптимально.

Если выбирать весовые коэффициенты k_j при линейном объединении ветвей так, чтобы они учитывали фактическое состояние каждой ветви, определяемое величиной h_{xj}, то при определенном правиле такого выбора можно добиться максимального отношения сигнал/шум на выходе устройства объединения. Поскольку величина h_{xj} в каждой ветви случайна и медленно изменяется во времени, необходимо обеспечить такое изменение весовых коэффициентов k_j для всех ветвей, чтобы величина $h_{xр}$ на выходе схемы объединения каждый раз достигала своего максимального значения. Подобное объедине-

ние ветвей называется линейным сложением при максимальном отношении сигнал/шум на выходе устройства объединения, или оптимальным линейным сложением.

Выясним, при каких условиях достигается оптимальное линейное сложение. Для того чтобы сложение было линейным, необходимо, как и в предыдущем случае, обеспечить фазирование всех копий сигнала в ветвях. Тогда результирующее колебание на выходе устройства объединения можно представить так.

$$y_{р}(t) = f_s(t) \sum_{j=1}^{n} k_j S_{xj} + \sum_{j=1}^{n} k_j n_j(t). \qquad (5.137)$$

Мощность сигнальной части колебания $y_р(t)$ определяется выражением

$$P_{cxр} = \frac{1}{\tau_0} \int_0^{\tau_0} s^2{}_{xр}(t)\, dt = \frac{1}{\tau_0} \int_0^{\tau_0} f^2{}_s(t) \left[\sum_{j=1}^{n} k_j S_{xj}\right]^2 dt. \qquad (5.138)$$

Так как случайные величины k_j и S_{xj} сохраняют на интервале 0—τ_0 постоянное значение, то

$$P_{cxр} = \left[\sum_{j=1}^{n} k_j S_{xj}\right]^2 \frac{1}{\tau_0} \int_0^{\tau_0} f^2{}_s(t)\, dt = \frac{1}{2} \left[\sum_{j=1}^{n} k_j S_{xj}\right]^2. \qquad (5.139)$$

Мощность шумовой части колебания $y_р(t)$ определяется выражением

$$P_{шр} = \frac{1}{\tau_0} \int_0^{\tau_0} \left[\sum_{j=1}^{n} k_j n_j(t)\right]^2 dt. \qquad (5.140)$$

Учтем, что случайная величина k_j и шум в ветви независимы, а шумы в отдельных ветвях некоррелированы. Тогда

$$P_{шр} = \sum_{j=1}^{n} k^2{}_j \sigma^2{}_ш, \qquad (5.141)$$

где $\sigma^2{}_ш$ — средний квадрат шума в ветви (мощность шума в ветви).

Отношение мощности сигнала к мощности шума на выходе схемы объединения в соответствии с (5.139) и (5.141) определяется выражением

$$h^2{}_{xр} = P_{cxр}/P_{шр} = 0{,}5 \left(\sum_{j=1}^{n} k_j S_{xj}\right)^2 \Bigg/ \sum_{j=1}^{n} k^2{}_j \sigma^2{}_ш. \qquad (5.142)$$

В соответствии с неравенством Буняковского — Шварца [5]

$$\left(\sum_{j=1}^{n} k_j S_{xj}\right)^2 \leqslant \left(\sum_{j=1}^{n} k^2{}_j\right)\left(\sum_{j=1}^{n} S^2{}_{xj}\right).$$

Тогда

$$h^2{}_{x\text{р}} \leqslant 0{,}5 \sum_{j=1}^{n} S^2{}_{xj}/\sigma^2{}_{\text{ш}} = \sum_{j=1}^{n} h^2{}_{xj}. \qquad (5.143)$$

Выражение (5.143) показывает, что при линейном сложении ветвей с разными весами отношение мощности сигнала к мощности шума на выходе схемы объединения не может превысить сумму соответствующих отношений в ветвях.

Условие, при котором выражение (5.143) переходит в равенство, соответствует оптимальному линейному сложению. Найдем его. Заменим величину $k^2{}_j\sigma^2{}_{\text{ш}}$ в выражении (5.142) эквивалентной величиной

$$k^2{}_j\sigma^2{}_{\text{ш}} = \beta^2 h^2{}_{xj}, \qquad (5.144)$$

в которой β — постоянный коэффициент, не зависящий от номера ветви. Тогда выражение (5.142) примет вид

$$h^2{}_{x\text{р}} = \left(\sum_{j=1}^{n} h_{xj} \frac{S_{xj}}{\sqrt{2}\,\sigma_{\text{ш}}}\right)^2 \Bigg/ \sum_{j=1}^{n} h^2{}_{xj}. \qquad (5.145)$$

Так как

$$S_{xj}/\sqrt{2}\,\sigma_{\text{ш}} = h_{xj},$$

то выражение (5.145) равно

$$h^2{}_{x\text{р}} = \sum_{j=1}^{n} h^2{}_{xj}, \qquad (5.146)$$

что соответствует оптимальному линейному сложению. Условием такого сложения является выбор весовых коэффициентов k_j в соответствии с выражением (5.144):

$$k_j = \beta \frac{h_{xj}}{\sigma_{\text{ш}}} = \beta \frac{S_{xj}}{\sqrt{2}\,\sigma^2{}_{\text{ш}}} = \beta\mu_j \frac{S_0}{\sqrt{2}\,\sigma^2{}_{\text{ш}}}. \qquad (5.147)$$

Из этого выражения следует, что значения весовых коэффициентов k_j должны быть пропорциональны случайной величине S_{xj} или μ_j (поскольку все остальные величины постоянны).

Таким образом, при оптимальном объединении ветвей копии складываются когерентно и с тем большим весом, чем выше уровень их огибающей. Для реализации этого условия необходимо в каждой ветви выполнить фазировку копий, а также обеспечить автоматическую регулировку усиления, при которой сильным замираниям сигнала (малым значениям S_{xj} или μ_j) соответствует малый коэффициент усиления, а слабым (большим значениям S_{xj} или μ_j) — большой. Иными словами, необходима такая регулировка усиления, режим работы которой противоположен режиму работы обычной АРУ, применяемой в приемниках.

Необходимость фазировки копий и соответствующей регулировки усиления в ветвях означает, что в приемном устройстве должна обеспечиваться точная оценка (измерение) случайных значений параметров принимаемого сигнала (фазы и уровня) в каждой ветви.

Рассмотрение выражения (5.146) позволяет выяснить вопрос с величине энергетического выигрыша и законе распределения случайной величины h_{xp}. Действительно, усреднив выражение (5.146), получим

$$\overline{h^2_{xp}} = \sum_{j=1}^{n} \overline{h^2_{xj}}. \tag{5.148}$$

При одинаковом значении $\overline{h^2_{xj}}$ во всех ветвях, равном h^2, имеем

$$\overline{h^2_{xp}} = nh^2; \tag{5.149}$$

отсюда следует

$$B_{\text{опт}} = n. \tag{5.150}$$

Это выражение определяет энергетический выигрыш при оптимальном линейном сложении ветвей. Величина выигрыша при таком объединении ветвей не зависит от статистики замираний сигнала в отдельных ветвях и определяется только числом ветвей.

В теории разнесенного приема такой выигрыш иногда называют бреннановским в честь американского ученого Бреннана, внесшего заметный вклад в эту теорию и впервые исследовавшего рассмотренную выше задачу [5]. Из выражения (5.146) следует, что закон распределения случайной величины h^2_{xp} определяется композицией законов распределения случайных величин h^2_{xj}. Если замирания в ветвях подчиняются закону Релея,

нетрудно показать, что распределение каждой величины $h^2{}_{xj}$ имеет вид

$$W(h^2{}_{xj}) = \frac{1}{h^2}\exp\left(-\frac{h^2{}_{xj}}{h^2}\right). \qquad (5.151)$$

Это экспоненциальное распределение. Его можно получить, если перейти от распределения величины h_{xj} к распределению ее квадрата $h^2{}_{xj}$. Для этого нужно учесть выражение (5.64) и выполнить преобразование в соответствии с формулой (5.58).

В теории вероятностей доказано, что плотность вероятности суммы экспоненциально распределенных случайных величин имеет так называемое χ^2-распределение [5, 11, 12]. Если все ветви независимы и одинаковы, это распределение определяется выражением

$$W(h^2{}_{xp}) = \frac{1}{(n-1)!}\frac{1}{h^2}\left(\frac{h^2{}_{xp}}{h^2}\right)^{n-1}\exp\left(-\frac{h^2{}_{xp}}{h^2}\right). \qquad (5.152)$$

От распределения (5.152) несложно перейти к распределению $W(h_{xp})$, которое имеет вид

$$W(h_{xp}) = \frac{2h_{xp}}{(n-1)!\,h^2}\left(\frac{h^2{}_{xp}}{h^2}\right)^{n-1}\exp\left(-\frac{h^2{}_{xp}}{h^2}\right). \qquad (5.153)$$

При $n=1$ это выражение совпадает с законом Релея, а при $n\to\infty$ стремится к δ-функции. Иными словами, при увеличении числа ветвей разнесения флюктуации сигнала на выходе схемы объединения уменьшаются.

Найдем вероятность ошибки рассматриваемого способа разнесенного приема. Будем полагать, как и ранее, что анализ колебания $y_p(t)$ выполняется некогерентным способом. Приняв во внимание (5.76), (5.153) и (5.109), имеем

$$\overline{p_{ош}} = \frac{1}{(n-1)!}\int_0^\infty \frac{h_{xp}}{h^2}\left(\frac{h^2{}_{xp}}{h^2}\right)^{n-1}\exp\left[-\frac{h^2{}_{xp}}{h^2}\times\right.$$

$$\left.\times\left(1+\frac{\gamma^2{}_c h^2}{2}\right)\right]dh_{xp}. \qquad (5.154)$$

После замены переменной этот интеграл примет вид табличного интеграла [22]:

$$\overline{p_{ош}} = \frac{1}{2(n-1)!}\int_0^\infty x^{n-1}\exp(-ax)\,dx = \frac{1}{2a^n}. \qquad (5.155)$$

С учетом (5.155) имеем

$$\overline{p_{\text{ош}}} = 0,5\,(1 + 0,5\gamma^2_{\text{c}} h^2)^{-n}. \qquad (6.156)$$

При большом отношении сигнал/шум в ветвях ($h^2 \gg 1$) формула упрощается и принимает вид

$$\overline{p_{\text{ош}}} \approx 0,5\,(2/\gamma^2_{\text{c}} h^2)^{-n}. \qquad (5.157)$$

Формула (5.157) проста и удобна для получения оценок эффективности разнесенного приема, выяснения влияния числа ветвей разнесения на помехоустойчивость приема и т. п.

Сопоставим теперь все три рассмотренных метода и обсудим полученные результаты.

5.7.4. Сравнение методов разнесенного приема с линейным объединением ветвей

Сравнение методов разнесенного приема начнем с рассмотрения величины энергетического выигрыша, зависимость которого от числа ветвей для трех методов показана на рис. 5.11. Из этого рисунка следует, что наиболее эффективен с энергетической точки зрения метод с оптимальным линейным объединением ветвей, а наименее эффективен — метод с автовыбором лучшей ветви. Метод с линейным сложением ветвей несколько уступает оптимальному.

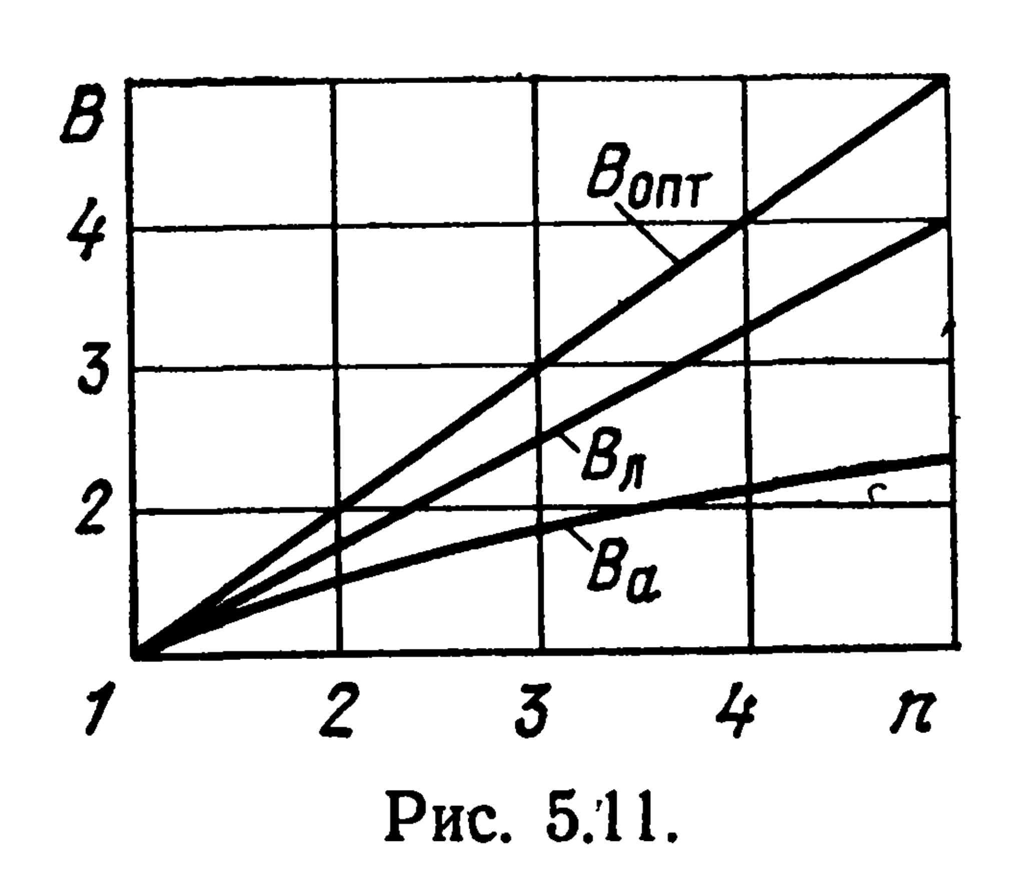


Рис. 5.11.

Величина выигрыша не может служить исчерпывающей характеристикой метода разнесенного приема и отражает только одну из его особенностей. Для более полной характеристики методов разнесенного приема необходимо располагать законами распределения огибающей сигнала на выходе схемы объединения. На рис. 5.12 представлены законы распределения на выходе схемы объединения при автовыборе и оптимальном линейном сложении для $n=2$ и $n=3$. На этом же рисунке приведен закон Релея, соответствующий одиночному приему флюктуирующего сигнала ($n=1$). Кривые построены

для обобщенных законов распределения, которые определяются выражением

$$W(x) = nx \exp(-x^2/2)[1 - \exp(-x^2/2)]^{n-1} \qquad (5.158)$$

для автовыбора и выражением

$$W(x) = (x/(n-1)!)(x^2/2)^{n-1} \exp(-x^2/2) \qquad (5.159)$$

для оптимального линейного сложения.

Выражения (5.158) и (5.159) соответствуют выражениям (5.120) и (5.153), в которых проведена следующая замена переменной: $x = \sqrt{2}\, h_{xp}/h$.

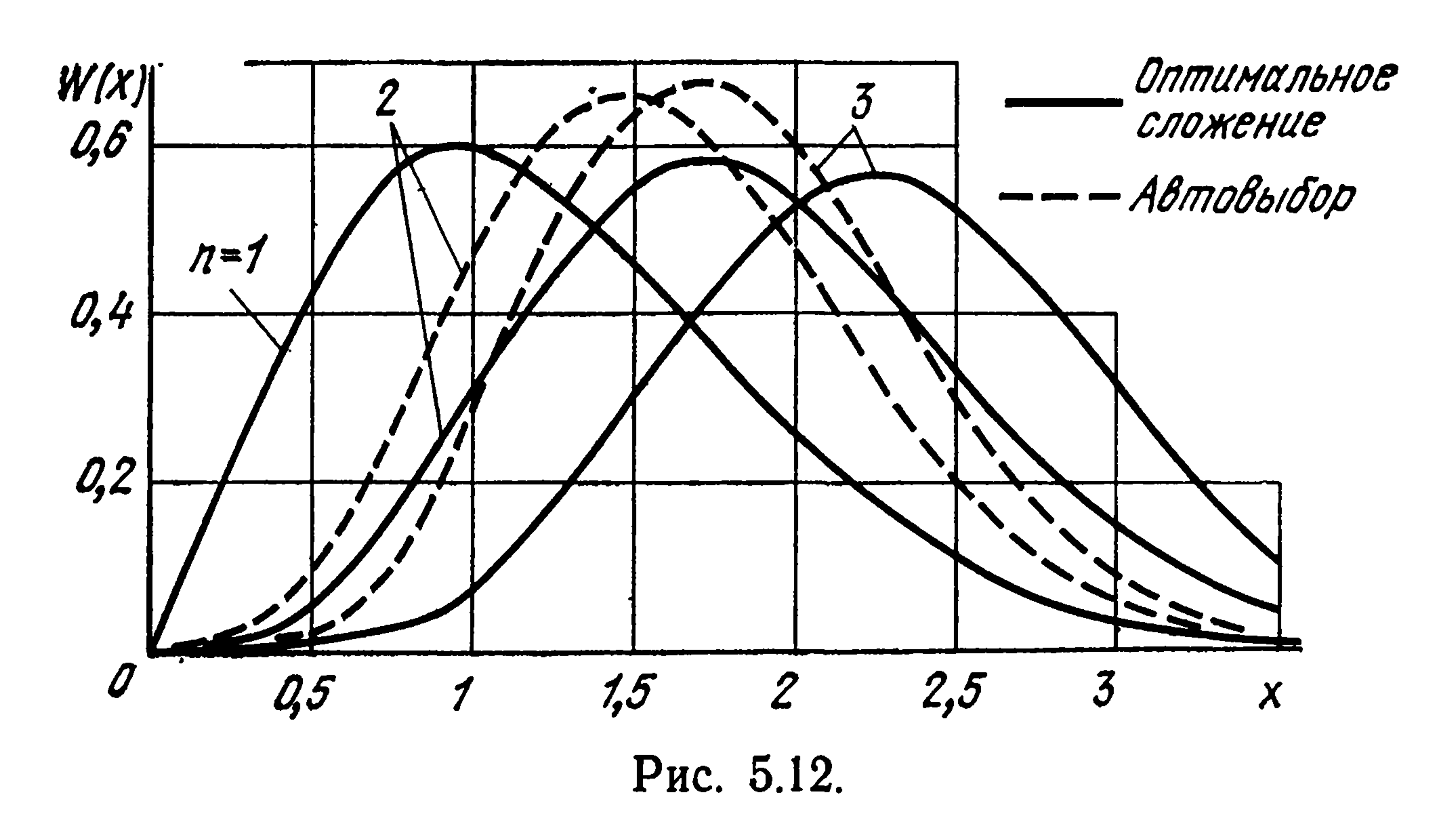


Рис. 5.12.

Из рассмотрения кривых (рис. 5.12) следует, что увеличение числа ветвей разнесения изменяет закон распределения в более «благоприятную» сторону — распределение смещается вправо, его дисперсия уменьшается, а форма приближается к виду нормального распределения. Из кривых также видно, что этот процесс проходит более эффективно для оптимального линейного сложения.

Как уже указывалось ранее, для линейного сложения ветвей получить распределение огибающей колебания на выходе устройства объединения ветвей затруднительно. Однако на основании проведенного ранее рассмотрения можно утверждать, что это распределение должно занимать промежуточное положение между распределениями для автовыбора и для оптимального линейного сложения.

Наиболее полной характеристикой метода разнесенного приема является его помехоустойчивость, определяемая вероятностью ошибки приема. В качестве крите-

рия сравнения помехоустойчивости методов удобно взять среднее значение отношения мощности сигнала к шуму в отдельной ветви, при котором обеспечивается заданная вероятность ошибки. Используя выражения (5.126) и (5.157), нетрудно показать, что отношение средних мощностей сигнала при автовыборе и оптимальном сложении (при одинаковых вероятностях ошибки) определяется выражением

$$P_{с\ а}/P_{с\ опт} = h^2{}_{а}/h^2{}_{опт} = (n!)^{1/n}. \qquad (5.160)$$

Расчеты показывают, что при увеличении числа ветвей автовыбор начинает заметно проигрывать по помехоустойчивости оптимальному сложению. Так, проигрыш составляет 1,6 дБ при $n=2$; 2,6 дБ при $n=3$ и 3,4 дБ при $n=4$.

Как уже указывалось, получить формулу для вероятности ошибки при линейном объединении ветвей не удается. Расчеты этой вероятности, выполненные на ЭВМ для различных значений n, приведены в работе [7]. Эти результаты показывают, что проигрыш метода линейного сложения по сравнению с оптимальным при любой кратности разнесения не превышает 1 дБ.

Сравнение рассмотренных методов разнесенного приема показывает, что несмотря на некоторое их различие в помехоустойчивости все они эффективны по отношению к одиночному приему.

Чтобы проследить зависимость помехоустойчивости разнесенного приема от энергетического отношения сигнал/шум в отдельной ветви, а также от числа ветвей разнесения (кратности разнесения), на рис. 5.13 приведены кривые, построенные по формуле (5.156) для некогерентного приема ФМн сигналов. На этом же рисунке приведены кривые для одиночного приема ($n=1$) в гауссовском и релеевском каналах. Из рассмотрения кривых рис. 5.13 следует, что уже при двух-трех ветвях разнесения дости-

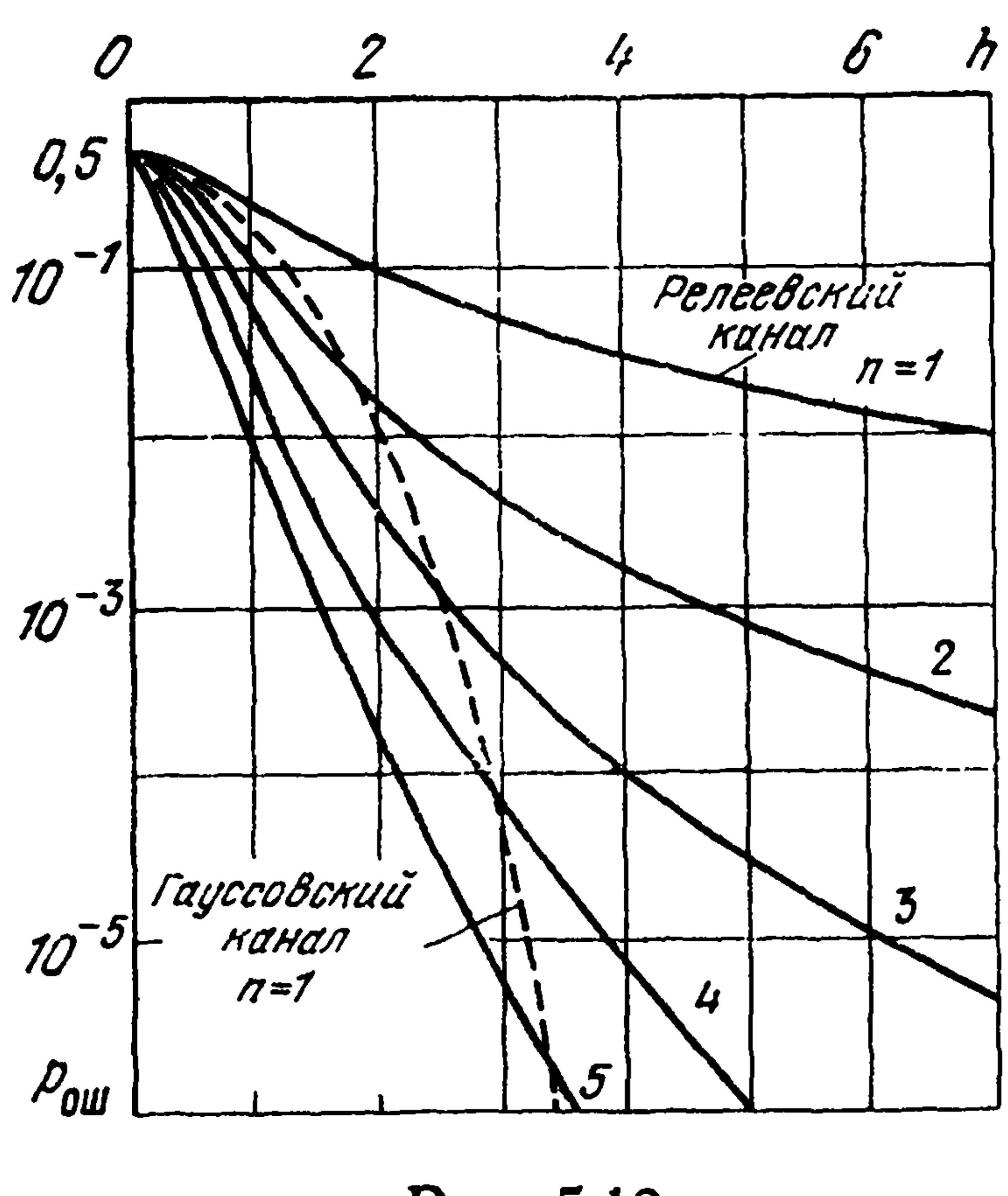


Рис. 5.13.

гается существенный выигрыш в помехоустойчивости по сравнению с одиночным приемом в релеевском канале. Дальнейшее увеличение кратности разнесения приводит к сравнительно медленному возрастанию этого выигрыша. Поэтому применять более чем две-три ветви разнесения практически часто оказывается нецелесообразно.

Из рассмотренных методов разнесенного приема наиболее удобен для практической реализации метод линейного объединения ветвей. Обеспечивая почти такие же показатели, как и метод оптимального сложения ветвей, он значительно проще в реализации. Когерентное сложение копий при линейном объединении ветвей достигается системой ФАП.

5.8. МЕТОДЫ РАЗНЕСЕННОГО ПРИЕМА С ПОСЛЕДЕТЕКТОРНЫМ ОБЪЕДИНЕНИЕМ ВЕТВЕЙ

Методы разнесенного приема, основанные на додетекторном объединении ветвей, предполагают, что все копии сигнала поступают на вход схемы объединения одновременно. Это условие означает отказ от учета времени многолучевого растяжения отдельных копий сигнала при анализе вопросов приема информации в каналах со случайными параметрами. На практике часто время растяжения составляет заметную долю от длительности посылки и пренебрегать этим явлением уже нельзя.

Другим обстоятельством, затрудняющим применение указанных методов, является необходимость достаточно точной оценки параметров флюктуирующего сигнала в каждой ветви. Если ориентироваться на наиболее удобный для реализации метод линейного объединения ветвей, то для него нужно получать оценки фаз копий сигнала и осуществлять фазирование этих копий. Однако в ряде случаев условия работы в реальных каналах таковы, что подобная оценка на практике невозможна (например, если на некоторых интервалах время корреляции флюктуаций сигнала заметно уменьшается и нарушается условие $\tau_к \gg \tau_0$, что приводит к резкому увеличению погрешности оценки фазы).

В настоящее время известен ряд методов разнесенного приема, основанных на последетекторном объедине-

нии ветвей. Такие методы не требуют получения оценок параметров флюктуирующих сигналов и поэтому часто более удобны для практического применения. Рассмотрим два таких метода.

5.8.1. Разнесенный прием с некогерентным объединением ветвей

Особенностью этого метода является то, что в каждой ветви разнесения осуществляется анализ принятой копии сигнала точно так же, как и при одиночном некогерентном приеме, с той лишь разницей, что решение о переданном символе принимается не в ветвях, а после их объединения.

При двоичной передаче копия флюктуирующего сигнала на входе любой ветви разнесения соответствует одному из двух возможных переданных сигналов s_1 или s_2. Как правило, при разнесенном приеме используют

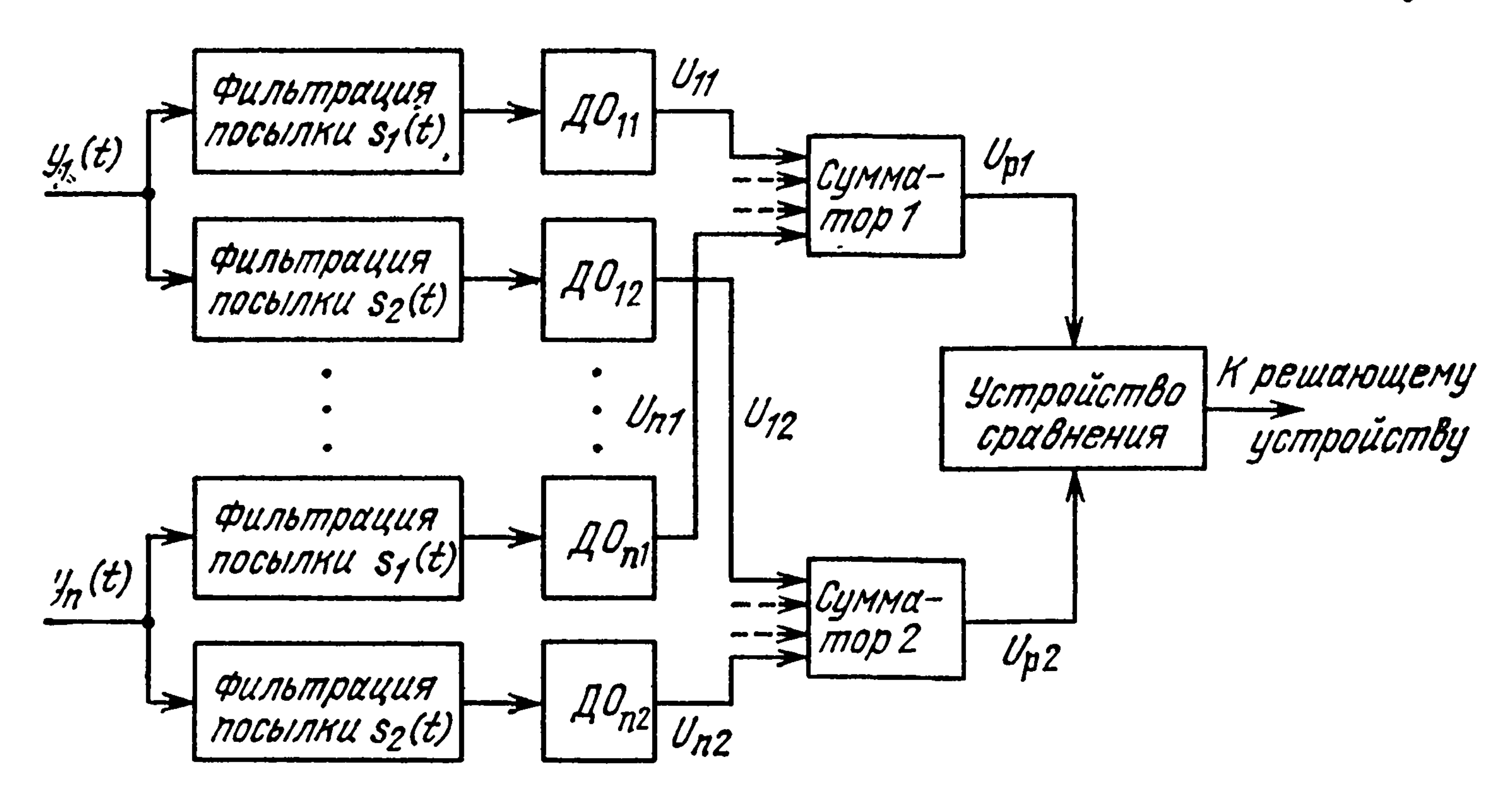


Рис. 5.14.

сигналы с активной паузой и одинаковыми энергиями посылок. Структурная схема разнесенного приема с некогерентным объединением ветвей, соответствующая такому условию, приведена на рис. 5.14. Колебание $y_j(t)$ на входе каждой ветви представляет собой сумму флюктуирующей копии сигнала $s_{xj}(t)$ и внутреннего шума $n_j(t)$. Анализ такого колебания выполняется фильтрацией принятой «зашумленной» копии сигнала с последующим выделением огибающей отфильтрованной копии детектором огибающей (ДО). Указанный анализ соответ-

ствует некогерентному способу приема в ветвях. Этот прием можно осуществить либо оптимально (в соответствии с методикой, рассмотренной в § 5.3), либо неоптимально (в соответствии со способами, рассмотренными в гл. 4). Поскольку различие между оптимальным и подоптимальным (реальным) некогерентным одиночным приемом по помехоустойчивости невелико, практически удобнее применять подоптимальный некогерентный прием.

Достоинством некогерентного приема флюктуирующих сигналов в отдельных ветвях является то, что такой прием не требует использования априорных сведений о случайных изменениях параметров этих сигналов (при условии медленности этих изменений по отношению к длительности отдельной посылки сигнала). Это позволяет упростить приемные устройства ветвей без существенного ухудшения помехоустойчивости и обеспечить инвариантность их структуры и параметров к указанным изменениям.

Если замирания в ветвях независимы, а среднее отношение сигнал/шум в ветвях одинаково, то объединение ветвей осуществляется раздельным сложением огибающих, соответствующих посылке сигнала $s_1(t)$, и огибающих, соответствующих посылке сигнала $s_2(t)$. Полученные в сумматорах *1* и *2* колебания

$$U_{p1}=\sum_{j=1}^{n} U_{j1} \text{ и } U_{p2}=\sum_{j=1}^{n} U_{j2} \qquad (5.161)$$

подаются на устройство сравнения, а результат сравнения

$$\Delta U_{p}=U_{p1}-U_{p2}$$

поступает на вход решающего устройства, в котором принимается решение о переданном символе.

Исследования рассмотренного метода разнесенного приема показали, что оптимальным способом некогерентного объединения ветвей является квадратичное сложение огибающих [7, 16, 20]. Это означает, что напряжения на выходе детекторов огибающих U_{1j} и U_{2j} должны быть пропорциональны квадратам огибающих соответствующих «зашумленных» посылок:

$$U_{j1}\equiv Y^2_{j1},\ U_{j2}\equiv Y^2_{j2}, \qquad (5.162)$$

где Y_{j1} и Y_{j2} — огибающие «зашумленных» посылок $s_{j1}(t)$ и $s_{j2}(t)$, т. е. огибающие колебаний

$$y_{j1}(t) = s_{xj1}(t) + n_j(t) = Y_{j1} f_{j1}(t),$$
$$y_{j2}(t) = s_{xj2}(t) + n_j(t) = Y_{j2} f_{j2}(t).$$

В таком случае суммы (5.161) можно записать так:

$$U_{p1} = \sum_{j=1}^{n} Y^2_{j1}, \quad U_{p2} \equiv \sum_{j=1}^{n} Y^2_{j2}. \tag{5.163}$$

Величины U_{p1} и U_{p2} являются случайными. Обозначим законы их распределения через $W(U_{p1})$ и $W(U_{p2})$. Найдем выражение, определяющее вероятность ошибочного приема сигнала $s_2(t)$ вместо фактически посланного сигнала $s_1(t)$. Для этого можно воспользоваться методикой, изложенной в § 4.3, в соответствии с которой искомая вероятность определяется выражением

$$p(s_2 \mid s_1) = \int_0^\infty \left[\int_0^{U_{p1}} W(U_{p2})\, dU_{p2} \right] W(U_{p1})\, dU_{p1}. \tag{5.164}$$

При посылке сигнала s_1 напряжение U_{p1} будет определяться копиями этого сигнала в ветвях и шумами в них, а напряжение U_{p2} — только шумами в ветвях. Так как схема симметрична относительно посылок s_1 и s_2, выражение (5.164) определяет вероятность ошибки данного метода разнесения.

Чтобы конкретизировать выражение (5.164), необходимо знать распределения $W(U_{p1})$ и $W(U_{p2})$. Если замирания сигнала в ветвях подчиняются релеевскому закону распределения, то на основании рассмотрения, проведенного в п. 5.7.3 можно утверждать, что величина U_{p1} имеет χ^2-распределение. Нетрудно видеть, что такое же распределение имеет и величина U_{p2}, определяемая суммой квадратов огибающих шума.

Вычисление вероятности (5.164) с учетом указанных распределений является громоздкой операцией, и мы этим заниматься не будем. Подобное вычисление выполнено в работах [16, 20]. Используя результаты [16], можно привести следующую формулу для вероятности ошибки при разнесенном приеме двоичных сигналов с некогерентным квадратичным объединением ветвей:

$$\overline{p_{ош}} = \sum_{k=0}^{n-1} C^k_{n+k-1} p^n_1 (1 - p_1)^k, \tag{5.165}$$

где n — число ветвей разнесения; p_1 — вероятность ошибки при одиночном некогерентном приеме в ветви с релеевскими замираниями.

Учитывая выражение (5.79), определяющее вероятность p_1, формулу (5.165) можно записать в виде

$$\overline{p_{\text{ош}}} = \frac{1}{(2+\gamma^2{}_{\text{c}}h^2)^n} \sum_{k=0}^{n-1} C^k{}_{n+k-1} \left(\frac{1+\gamma^2{}_{\text{c}}h^2}{2+\gamma^2{}_{\text{c}}h^2} \right)^k. \qquad (5.166)$$

Если $h \gg 1$ (например, $h \geqslant 10$), формула упрощается и с учетом известного в математике тождества [22]

$$\sum_{k=0}^{n-1} C^k{}_{n+k-1} = C^n{}_{2n-1},$$

принимает вид

$$\overline{p_{\text{ош}}} \approx \frac{C^n{}_{2n-1}}{(\gamma^2{}_{\text{c}}h^2)^n} = \frac{(2n-1)!}{n!\,(n-1)!} \frac{1}{(\gamma^2{}_{\text{c}}h^2)^n}. \qquad (5.167)$$

При $n=2$ и $n=3$ имеем соответственно

$$\overline{p_{\text{ош}}} = \frac{3}{4}\left(\frac{2}{\gamma^2{}_{\text{c}}h^2}\right)^2, \quad \overline{p_{\text{ош}}} = \frac{10}{8}\left(\frac{2}{\gamma^2{}_{\text{c}}h^2}\right)^3. \qquad (5.168)$$

Сопоставление этих результатов с выражением (5.157) показывает, что различие в помехоустойчивости между рассмотренным методом и методом с оптимальным линейным объединением ветвей невелико (около 1 дБ). При увеличении числа ветвей это различие возрастает, однако, как указывалось выше, на практике большое число ветвей не применяется. Поэтому разнесенный прием с квадратичным сложением ветвей является эффективным методом борьбы с замираниями сигналов.

При реализации этого метода часто квадратичное сложение заменяют линейным сложением огибающих. Это связано с тем, что получить напряжения U_{j1} и U_{j2}, пропорциональные соответствующим огибающим Y_{j1} и Y_{j2}, технически значительно проще, чем их квадратам. В этом случае

$$U_{\text{p1}} \equiv \sum_{j=1}^{n} Y_{j1}, \quad U_{\text{p2}} \equiv \sum_{j=1}^{n} Y_{j2}. \qquad (5.169)$$

Нетрудно видеть, что распределения $W(U_{\text{p1}})$ и $W(U_{\text{p2}})$ теперь уже определяются композицией законов распределения составляющих сумм. При релеевских замираниях сигнала распределения $W(U_{\text{p1}})$ и $W(U_{\text{p2}})$

определяются композицией законов Релея. Следовательно, переход от квадратичного последетекторного объединения ветвей к линейному аналогичен переходу от оптимального когерентного сложения ветвей (взвешенного линейного сложения) к когерентному сложению с равными весами (невзвешенному линейному сложению). Это обстоятельство дает основание предполагать, что некогерентное последетекторное линейное сложение ветвей не должно существенно ухудшать помехоустойчивость приема по сравнению с квадратичным сложением. К сожалению, исследование помехоустойчивости при некогерентном линейном сложении ветвей наталкивается на такие же математические трудности, что и при когерентном линейном сложении.

В явном виде решить эту задачу удается только для сдвоенного приема [16]. Основываясь на результатах этой работы, несложно получить следующее выражение для вероятности ошибки при достаточно большом среднем отношении сигнал/шум в отдельной ветви ($h \gg 1$):

$$\overline{p_{\text{ош}}} \approx 0{,}8\,(2/\gamma^2_{\text{c}}h^2)^2, \quad n = 2. \qquad (5.170)$$

Практически это выражение совпадает с соответствующим выражением для квадратичного сложения (см. (5.168)). Нет оснований ожидать, что при числе ветвей разнесения $n=3$ или $n=4$ различие в помехоустойчивости станет значительным. Таким образом, при реализации разнесенного приема с последетекторным объединением ветвей целесообразно применять более простую схему линейного сложения, которая обеспечивает почти такую же помехоустойчивость, что и схема квадратичного сложения.

5.8.2. Разнесенный прием с дискретным сложением

Особенностью этого метода приема является то, что решение о переданном символе принимается в два этапа. На первом, предварительном этапе в каждой ветви принимается решение о том, какой информационный символ содержит данная флюктуирующая копия переданной посылки. Это означает, что в каждой ветви выполняются такие же операции, как при одиночном приеме. На втором, заключительном этапе на основании анализа отдельных решений, полученных в каждой ветви, прини-

мается окончательное решение о переданном символе. Если все ветви одинаковы, целесообразно принять решение о том символе, который зафиксирован в наибольшем числе ветвей. Это правило называют иногда голосованием по большинству или мажоритарным правилом. Очевидно, что при таком правиле для исключения неопределенности принятия решения при четном основании передаваемых цифровых последовательностей всегда необходимо использовать нечетное число ветвей разнесения. Следовательно, для разнесенного приема двоичных цифровых последовательностей необходимо иметь $n=(2k-1)$ ветвей, где $k=1, 2, \ldots, 3$. Практически, с учетом сложности реализации приема, речь может идти только о трех или (в редких случаях) пяти ветвях разнесения.

Используя результаты, полученные в работе [16], можно указать следующую формулу для оценки вероятности ошибки при разнесенном приеме с дискретным сложением:

$$\overline{p_{\text{ош}}}=\frac{(2k-1)!}{k!\,(k-1)!}\bar{p}^{k}{}_{1}, \qquad \bar{p}_1\leqslant 0,1, \tag{5.171}$$

где $\bar{p}_1$ — средняя вероятность ошибки приема в отдельной ветви.

При некогерентных способах приема в ветвях величина $\bar{p}_1$ определяется выражением (5.79), с учетом которого имеем

$$\overline{p_{\text{ош}}}\approx\frac{(2k-1)!}{k!\,(k-1)!}\left(\frac{1}{2+\gamma^2{}_{c}h^2}\right)^{k}. \tag{5.172}$$

Для трех ветвей ($n=3$, $k=2$)

$$\overline{p_{\text{ош}}}=3\overline{p^2{}_1}=3\,(2+\gamma^2{}_{c}h^2)^{-2} \tag{5.173}$$

Для пяти ветвей ($n=5$, $k=3$)

$$\overline{p_{\text{ош}}}=10\overline{p^3{}_1}=10\,(2+\gamma^2{}_{c}h^2)^{-3}. \tag{5.174}$$

При $h\gg 1$ эти выражения совпадают с выражениями (5.168), т. е. при трех ветвях дискретное сложение обеспечивает такую же помехоустойчивость, как сдвоенный прием с квадратичным сложением ветвей, а при пяти ветвях — как строенный прием.

Необходимость увеличения числа ветвей по сравнению с рассмотренными ранее методами приема может быть оправдана простотой реализации разнесенного приема с дискретным сложением. Для этого достаточно иметь

и одиночных приемников и дополнительное устройств объединения и «подсчета» частных решений с целью принятия окончательного решения.

Рассмотренные методы разнесенного приема с последетекторным объединением не требуют оценки (измерения) параметров флюктуирующих сигналов, сравнительно просты в реализации и обеспечивают помехоустойчивость, близкую к потенциальной. По вопросам разнесенного приема имеется обширная литература. Наиболее детальное рассмотрение этих вопросов содержится в работах [5, 16, 20, 21, 23].

5.9. КРАТКИЕ ИТОГИ ГЛ. 5

1. В этой главе были рассмотрены вопросы приема двоичных сигналов в каналах со случайными параметрами. При исследовании таких вопросов важное значение имеет выбор модели среды распространения сигналов. Эта модель должна учитывать основные особенности среды и в то же время быть достаточно простой и удобной для исследования.

2. В настоящее время весьма эффективным методом исследования многих реальных каналов со случайными параметрами является линейная трактовка среды распространения, в соответствии с которой среда рассматривается как линейный фильтр со случайно изменяющимися параметрами.

3. Для многих важных с прикладной точки зрения каналов связи указанный подход позволяет рассматривать среду как фильтр, случайный коэффициент передачи которого подчиняется обобщенному закону Релея и в некоторой полосе частот не зависит от частоты.

Сигнал, прошедший через такую среду, становится флюктуирующим, т. е. приобретает случайные изменения уровня и начальной фазы. При указанных выше условиях замирания уровня сигнала не приводят к изменению формы посылок сигнала. Такие замирания называются общими или мультипликативными.

4. Исследование вопросов одиночного приема двоичных сигналов с общими замираниями показывает, что помехоустойчивость такого приема оказывается низкой (особенно, если замирания подчиняются закону Релея, т. е. являются глубокими).

5. Эффективным методом повышения помехоустойчивости приема замирающих сигналов является разнесенный прием, основанный на обработке нескольких независимых копий сигнала, содержащих одну и ту же информацию.

6. Наиболее распространенными видами разнесения, обеспечивающими получение необходимых копий сигнала, являются угловое и пространственное разнесение.

7. Все известные методы разнесенного приема, использующие тот или иной вид разнесения сигналов, могут быть разделены на две группы: 1) разнесенный прием с додетекторным объединением копий сигнала; 2) разнесенный прием с последетекторным объединением копий сигнала.

8. Исследование различных методов разнесенного приема показывает, что все они обеспечивают существенное повышение помехоустойчивости по сравнению с одиночным приемом. Это повышение определяется не столько самим методом разнесенного приема, сколько числом ветвей разнесения.

9. Анализ показывает, что наибольший «прирост» помехоустойчивости обеспечивают 2—4 ветви разнесения. Дальнейшее увеличение числа ветвей разнесения увеличивает «прирост» помехоустойчивости в меньшей степени. Поэтому практически наиболее часто применяют не более двух-трех ветвей разнесения.

10. Наибольший интерес для практического применения разнесенного приема из первой группы методов (см. п. 7) представляет метод линейного (когерентного) объединения ветвей, а из второй группы — метод последетекторного линейного объединения огибающих копий сигнала. Эти методы сравнительно просты в реализации и обеспечивают помехоустойчивость, близкую к оптимальному разнесенному приему (при числе ветвей $n \leqslant 4$).

11. Разнесенный прием является одним из характерных примеров принципа дублирования. По сути дела, существенное повышение качества приема дискретной информации при разнесенном приеме обеспечивается только благодаря увеличению аппаратурной избыточности. Иными словами, повышение качества приема информации достигается увеличением аппаратурных «затрат» без изменения структуры первичных кодов, с помощью которых представлена цифровая информация, и без измене-

ния структуры сигналов, с помощью которых эта информация передается по каналу связи.

Известны методы, которые позволяют повысить качество приема информации в каналах со случайными параметрами не только вводя аппаратурную избыточность, но и изменяя структуру первичных кодов и структуру сигналов введением в них некоторой избыточности. С некоторыми из таких методов мы познакомимся в гл. 6.

КОНТРОЛЬНЫЕ ВОПРОСЫ И ЗАДАЧИ

5.1. Привести примеры каналов со случайными параметрами. Дать качественное объяснение причин случайного изменения параметров этих каналов.

5.2. Показать, что в упрощенной модели среды первый четырехполюсник, характеризующий идеальную среду, можно полностью описать коэффициентом ослабления мощности сигнала

$$k_р = GS_э/4\pi r^2$$

и временем запаздывания сигнала

$$\tau_{з0} = r/c_0,$$

где r — расстояние от передатчика до приемника; c_0 — скорость распространения колебаний в идеальной (однородной) среде данной физической природы; G — коэффициент направленного действия передающей антенны; $S_э$ — эффективная поверхность приемной антенны.

5.3. Объяснить, почему четырехполюсник, рассматриваемый в задаче 5.2, является неискажающим? Какова его импульсная переходная характеристика?

5.4. Показать, что если средняя величина коэффициента затухания сигнала, определяемая поглощением его энергии в среде, выражается соотношением

$$\alpha = kf^2 \text{ [дБ/км]},$$

где k — постоянный коэффициент, величина которого определяется свойствами конкретной среды, то частотная характеристика среды определяется выражением

$$K(\omega) = \exp(-3k\omega^2 r \cdot 10^{-3}).$$

5.5. Объяснить, почему время многолучевого растяжения зависит от вида диаграмм направленности передающей и приемной антенн и протяженности линии связи. Нарисовать примерный вид этой зависимости от указанных факторов.

5.6. Оценить возможную величину технической скорости передачи в каналах, ориентировочные значения времени многолучевого растяжения для которых приведены в § 5.2.

5.7. Каким коэффициентом растяжения должен обладать канал, если допустимо растяжение по времени до 10% и по частоте до 5%?

5.8. Линия связи использует многолучевой канал с максимальной разностью хода лучей $\Delta r_{max}=100$ м; скорость движения неоднородностей $U_{отр}=3$ м/с. Какую скорость передачи можно обеспечить по такой линии связи, если применять простые двоичные сигналы?

5.9. Почему применение автоматической регулировки усиления при одиночном приеме флюктуирующих сигналов не уменьшает среднюю вероятность ошибки приема?

5.10. Показать, что выражение (5.26) при увеличении β в пределе стремится к δ-функции.

Указание. При рассмотрении воспользоваться разложением вида

$$I_0(x)=\frac{\exp(x)}{\sqrt{2\pi x}}\left(1+\frac{1}{8x}+\ldots\right),\quad x>1,$$

а также учесть, что при $\beta\longrightarrow\infty$ с вероятностью, близкой к единице, $S_x/S_р\longrightarrow 1$.

5.11. Показать, что, если распределение огибающей посылок определяется выражением (5.25), то распределение мультипликативного коэффициента μ имеет вид

$$W(\mu)=\frac{2\mu}{\overline{\mu^2_{фл}}}\exp\left(-\frac{\mu^2+\mu^2_р}{\overline{\mu^2_{фл}}}\right)I_0\left(2\frac{\mu\mu_р}{\overline{\mu^2_{фл}}}\right),$$

где $\overline{\mu^2_{фл}}=\overline{\mu^2}-\mu^2_р$ — среднее значение квадрата флюктуирующей части коэффициента μ; $\mu_р$ — регулярная составляющая этого коэффициента.

5.12. Показать, что в райсовском канале закон распределения величины h_x определяется выражением

$$W(h_x)=\frac{2h_x}{h^2}\exp\left(-\frac{h^2_x}{h^2}-\frac{S^2_р}{2\sigma^2_{фл}}\right)I_0\left(\sqrt{2}\,\frac{S_р}{\sigma_{фл}}\frac{h_x}{h}\right).$$

5.13. Показать, что условную вероятность ошибки при некогерентных способах одиночного приема, определяемую выражением (5.76), можно записать в эквивалентном виде

$$p_{ош}(\mu)=0{,}5\exp\left(-\gamma^2_c\mu^2h^2/2\right),$$

где μ — коэффициент передачи среды со случайными параметрами.

5.14. Показать, что среднюю вероятность ошибки в райсовском канале с распределением коэффициента передачи среды в соответствии с выражением, приведенным в задаче 5.11, и условной вероятностью ошибки, определяемой формулой (5.76), можно выразить формулой

$$\overline{p}_{ош}=\frac{1}{2}\,\frac{A}{A+0{,}5\gamma^2_ch^2}\exp\left[-\frac{1}{2}\,\frac{(A-1)\gamma^2_ch^2}{A+0{,}5\gamma^2_ch^2}\right],$$

где

$$A=\left(\mu^2_р+\overline{\mu^2_{фл}}\right)/\overline{\mu^2_{фл}}.$$

Указание. Учесть

$$1)\ \overline{\mu^2} = \mu^2_р + \overline{\mu^2_{фл}} = 1,$$

$$\overline{\mu^2_{фл}} = \frac{\overline{\mu^2_{фл}}}{\overline{\mu^2}} = \frac{\overline{\mu^2_{фл}}}{\mu^2_р + \overline{\mu^2_{фл}}} = \frac{1}{A};$$

$$2)\ \int_0^\infty x \exp(-\alpha x^2) I_0(\beta x) dx = \frac{1}{2\alpha} \exp(\beta^2/4\alpha).$$

Решение подобной задачи имеется в [16].

5.15. Показать, что при $A \to 1$ формула для $\bar{p}_{ош}$, приведенная в предыдущей задаче, совпадает с выражением (5.79), а при $A \to \infty$ — с выражением (4.80).

5.16. Показать, что при некогерентном приеме двоичных равновероятных сигналов, передаваемых методом ОФМн, скорость передачи в релеевском канале уменьшается по сравнению с гауссовским в соответствии с выражением

$$R_Р/R_Г = h^2/\exp(h^2),$$

где h^2 — необходимое отношение сигнал/шум в гауссовском канале, при котором обеспечивается заданная вероятность ошибки.

Указание. При сравнении скоростей передач необходимо полагать вероятность ошибки в обоих случаях одинаковой и достаточно малой ($p_{ош} \ll 1$).

5.17. Дать сравнительную характеристику различных видов разнесения сигналов.

5.18. Объяснить, почему коэффициент корреляции между копиями сигналов при пространственном разнесении в вертикальном или горизонтальном направлении, перпендикулярном трассе распространения сигналов, падает быстрее, чем при разнесении вдоль трассы?

5.19. Изложить суть методов разнесенного приема с додетекторным и последетекторным объединением ветвей.

5.20. Указать основные недостатки метода автовыбора.

5.21. Показать, что при оптимальном линейном объединении ветвей огибающая результирующего колебания пропорциональна сумме квадратов огибающих копий в отдельных ветвях.

5.22. Объяснить, почему при оптимальном линейном объединении ветвей закон распределения случайной величины $h_{xр}$ отличается от закона распределения огибающей $S_{xр}$.

5.23. Показать, что при оптимальном линейном объединении ветвей случайная величина $h_{xр}$ распределена по закону (5.153).

5.24. Во сколько раз нужно увеличить мощность сигнала, чтобы при оптимальном линейном объединении ветвей двукратное разнесение стало эквивалентно по вероятности ошибки трехкратному?

5.25. Сравнить помехоустойчивость всех рассмотренных в гл. 5 методов разнесенного приема при следующих данных: $h^2 = 50$, $n = 3$, сигнал с относительной фазовой манипуляцией $\varphi_1 = 0$; $\varphi_2 = \pi$.

СПИСОК ЛИТЕРАТУРЫ

1. **Кириллов Н. Е.** Помехоустойчивая передача сообщений по линейным каналам со случайно изменяющимися параметрами. М., «Связь», 1971.

2. **Кеннеди Р.** Каналы связи с замираниями и рассеянием. Пер. с англ. М., «Сов. радио», 1973.

2. **Тарасюк Ю. Ф., Серавин Г. Н.** Гидроакустическая телеметрия. Л., «Судостроение», 1973.

4. **Каналы** передачи данных. Под ред. В. О. Шварцмана. М., «Связь», 1970.

5. **Лекции** по теории систем связи. Пер. с англ. М., «Мир», 1964.

6. **Долуханов М. П.** Флюктуационные процессы при распространении радиоволн. М., «Связь», 1971.

7. **Стейн С., Джонс Дж.** Принципы современной теории связи и их применение к передаче дискретных сообщений. М., «Связь», 1971.

8. **Принципы** отождествления каналов передачи сигналов. Под ред. Л. И. Филиппова. М., «Наука», 1973.

9. **Пенин П. И., Цвелев Э. А.** Упрощенная модель среды информационных каналов с частотно-зависимыми характеристиками и случайными параметрами. Труды МЭИ, 1972, вып. 117.

10. **Радиолинии** ионосферного рассеяния метровых волн. Под ред. Н. Н. Шумской. М., «Связь», 1973.

11. **Левин Б. Р.** Теоретические основы статистической радиотехники. Т. I. М., «Сов. радио», 1966.

12. **Тихонов В. И.** Статистическая радиотехника. М. «Сов. радио», 1966.

13. **Миддлтон Д.** Введение в статистическую теорию связи. Т. 2. Пер. с англ. М., «Сов. радио», 1962.

14. **Гуткин Л. С.** Теория оптимальных методов радиоприема при флюктуационных помехах. М., «Сов. радио», 1972.

15. **Ван Трис Г.** Теория обнаружения, оценок и модуляции. Пер. с англ. М., «Сов. радио», 1972.

16. **Финк Л. М.** Теория передачи дискретных сообщений. М., «Сов. радио», 1970.

17. **Филиппов Л. И.** Основы теории радиоприема дискретных сигналов. М., «Наука», 1974.

18. **Тейлор Р.** Моноимпульсная система сопровождения спутников, использующая поляризационно разнесенный прием. — «Зарубежная радиоэлектроника», 1968, № 2.

19. **Калинин А. И., Черенкова Е. Л.** Распространение радиоволн и работа радиолиний. М., «Связь», 1971.

20. **Андронов И. С., Финк Л. М.** Передача дискретных сообщений по параллельным каналам. М., «Сов. радио», 1971.

21. **Зюко А. Г.** Помехоустойчивость и эффективность систем связи. М., «Связь», 1972.

22. **Градштейн И. С., Рыжик И. М.** Таблицы интегралов, сумм, рядов и произведений. М., «Физматгиз», 1962.

23. **Кловский Д. Д.** Передача дискретных сообщений по радиоканалам. М., «Связь», 1969.

Глава 6

Методы повышения качества работы систем передачи цифровой информации

6.1. ВВЕДЕНИЕ

При создании системы передачи цифровой информации необходимо учитывать не только ее назначение, условия работы, характер решаемых задач и требуемые показатели качества передаваемой информации, но и ряд других факторов. К их числу относятся неизбежные физические ограничения, связанные с типом используемого канала, шириной выделенного диапазона рабочих частот, техническими возможностями реализации данной системы, экономическими соображениями и т. д.

В зависимости от требований, предъявляемых к системе, и конкретных условий ее работы, роль одних и тех же ограничений может быть различной. Например если в канале с ограниченной полосой пропускания необходимо обеспечить достаточно высокую скорость передачи информации при заданной вероятности ошибки, то существенное значение при выборе способа передачи и приема имеет ограничение по излучаемой пиковой мощности. Если в таком же канале нужно передавать информацию с очень малой вероятностью ошибки и не требовать высокой скорости (например, передача команд, передача сигналов синхронизации и т. п.), то ограничение по излучаемой пиковой мощности не столь важно, как в первом случае.

Как уже указывалось в гл. 1, полное решение задачи оптимизации системы передачи информации при ряде наложенных ограничений до сих пор не известно. Поэтому на практике обычно стремятся получить требуемые показатели только с учетом тех ограничений, которые наиболее существенны для рассматриваемой системы. Практически такой путь во многих случаях оправдан, так как показатели разумно спроектированных систем часто оказываются довольно близкими к показателям оптимальных систем.

В настоящее время известны следующие пути повышения качества работы систем передачи цифровой информации:

1) применение многопозиционных сигналов, эффективно использующих отведенную полосу частот; 2) применение сложных сигналов, эффективно использующих излучаемую энергию; 3) применение корректирующих кодов; 4) применение систем с обратным каналом связи. При создании систем часто комбинируют указанные выше возможности, что позволяет получить высокие качественные показатели работы систем.

Цель этой главы состоит в кратком изложении основных методов повышения качества работы систем и рассмотрении примеров, иллюстрирующих приложение этих методов к некоторым практическим задачам передачи цифровой информации.

6.2. МНОГОПОЗИЦИОННЫЕ СИГНАЛЫ, ЭФФЕКТИВНО ИСПОЛЬЗУЮЩИЕ ОТВЕДЕННУЮ ПОЛОСУ ЧАСТОТ

В ряде случаев канал связи, выделенный для передачи цифровой информации, обладает ограниченной полосой пропускания частот. Такое положение, например, имеет место при передаче данных по телефонным каналам, при передаче телеметрической информации в цифровой форме по гидроакустическим каналам и т. п. Во всех этих случаях желательно использовать предоставленную полосу частот наиболее эффективо. Это означает, что необходимо спроектировать систему связи с малыми (по возможности) удельными затратами полосы. В гл. 1 было показано, что для этого необходимо применять многоосновные (многоалфавитные) сигналы с минимально возможной базой. К их числу относятся сигналы с многопозиционной амплитудной или фазовой манипуляцией. При определенных условиях такие сигналы позволяют повысить скорость передачи информации без расширения занимаемой полосы частот. Выясним суть этих условий и оценим возможности применения указанных сигналов.

Если при формировании цифровых последовательностей используется алфавит из n символов, то для их передачи необходимо располагать таким же числом разли-

чающихся между собой посылок сигнала. Будем далее обозначать это число через M ($M=n$).

При многопозиционной АМн посылки сигнала различаются лишь амплитудами. Такой сигнал можно записать в виде

$$s_i(t)=\begin{cases} S_{0i}\sin(\omega_0 t+\varphi_0), & 0\leqslant t\leqslant\tau_0 \\ 0 & \text{для других } t, \end{cases} \tag{6.1}$$

где $i=1, 2, \ldots, M$; φ_0 — постоянная начальная фаза.

Полагая, что все символы (а следовательно, и посылки сигнала) равновероятны, а излучаемый сигнал ограничен по пиковой мощности, нетрудно убедиться, что амплитуды посылок определяются выражением

$$S_{0i}=\frac{S_0}{M}(i-1)=\Delta S_0(i-1). \tag{6.2}$$

При передаче информации многопозиционными АМн сигналами более вероятными ошибками под действием флюктуационной помехи будут ошибки, связанные с приемом посылок сигналов, близких к переданной. Например, при передаче посылки s_i ошибочный прием посылок s_{i-1} или s_{i+1} более вероятен, чем ошибочный прием посылок s_{i-2} или s_{i+2}. Это естественно, поскольку условия различения близких посылок хуже, чем посылок с заметно различающимися амплитудами.

Вероятность ошибки приема символа многопозиционного сигнала можно записать в виде

$$p_{\text{ош}}=\sum_{i=1}^{M}\sum_{\substack{j=1\\(i\neq j)}}^{M}p(s_i)\,p(s_j\,|\,s_i), \tag{6.3}$$

где $p(s_i)$ — априорная вероятность посылки с номером i; $p(s_j/s_i)$ — условная вероятность приема посылки s_j вместо переданной посылки s_i.

Так как

$$p(s_{i+1}\,|\,s_i)=p(s_{i-1}\,|\,s_i) \text{ и } p(s_{i+2}\,|\,s_i)\ll p(s_{i+1}\,|\,s_i),$$

то при равновероятных посылках выражение (6.3) с точностью, достаточной для инженерных приложений, можно заменить более простым приближенным выражением

$$p_{\text{ош}}\approx\frac{2(M-1)}{M}\,p(s_{i-1}\,|\,s_i). \tag{6.4}$$

При оптимальном приеме (см. гл. 3)

$$p(s_{i-1}|s_i)=0,5-\Phi(h_M/\sqrt{2}),$$

где в данном случае величина h_M зависит от энергии разности сигналов s_{i-1} и s_i. Основываясь на результатах гл. 3, нетрудно показать, что

$$h^2_M=\Delta S^2_0\tau_0/2N_0=S^2_0\tau_0/2M^2N_0. \qquad (6.5)$$

Поскольку каждая посылка многопозиционного сигнала содержит $\log_2 M$ двоичных единиц информации, то энергетические затраты на передачу одной двоичной единицы составляют

$$h^2_0=\frac{h^2_M}{\log_2 M}=\left(\frac{P_{\text{c max}}}{P_{\text{ш}}}\right)_{\text{вх}}\frac{1}{M^2\log_2 M}, \qquad (6.6)$$

где

$$P_{\text{c max}}=S^2_0\tau_0/2;\ P_{\text{ш}}=N_0\Delta f_{\text{э}};\ \tau_0\Delta f_{\text{э}}\approx 1.$$

Для того чтобы обеспечивалась заданная достоверность передачи информации, необходимо выполнить условие

$$(P_{\text{c max}}/P_{\text{ш}})_{\text{вх}}\geqslant h^2_{0\,\text{тр}}M^2\log_2 M, \qquad (6.7)$$

где $h^2_{0\,\text{тр}}$ — величина, при которой обеспечивается заданная вероятность ошибки.

Из выражения (6.7) следует, что по сравнению с двоичной АМн многопозиционная АМн требует существенного увеличения отношения сигнал/шум на входе приемника. Так, например, при $M=4$ необходимо увеличивать это отношение в 32 раза.

Рассмотрим теперь передачу цифровой информации многопозиционными ФМн сигналами. В этом случае все посылки имеют одинаковую амплитуду и частоту и различаются только начальными фазами. Такой сигнал можно записать в виде

$$s_i(t)=\begin{cases} S_0\sin(\omega_0 t+\varphi_{0i}), & 0\leqslant t\leqslant\tau_0,\\ 0 & \text{для других } t,\end{cases} \qquad (6.8)$$

где $i=1, 2, \ldots, M$.

Обычно при многопозиционной ФМн число фаз (а следовательно, и число посылок) берется кратным двум, т. е. $M=2^k$. Величину k часто называют кратностью фазовой манипуляции.

При равновероятных посылках их начальные фазы равны

$$\varphi_{0i} = \frac{2\pi}{M}(i-1) = \frac{2\pi}{2^k}(i-1) = \Delta\varphi_0(i-1). \qquad (6.9)$$

Геометрическое представление многократной ФМн показано на рис. 6.1. Если помех нет, то посылка сигнала соответствует концу вектора сигнала. Под действием помехи результирующий вектор может либо остаться в области данной посылки, либо перейти в любую другую область. Из рисунка следует, что наиболее вероят-

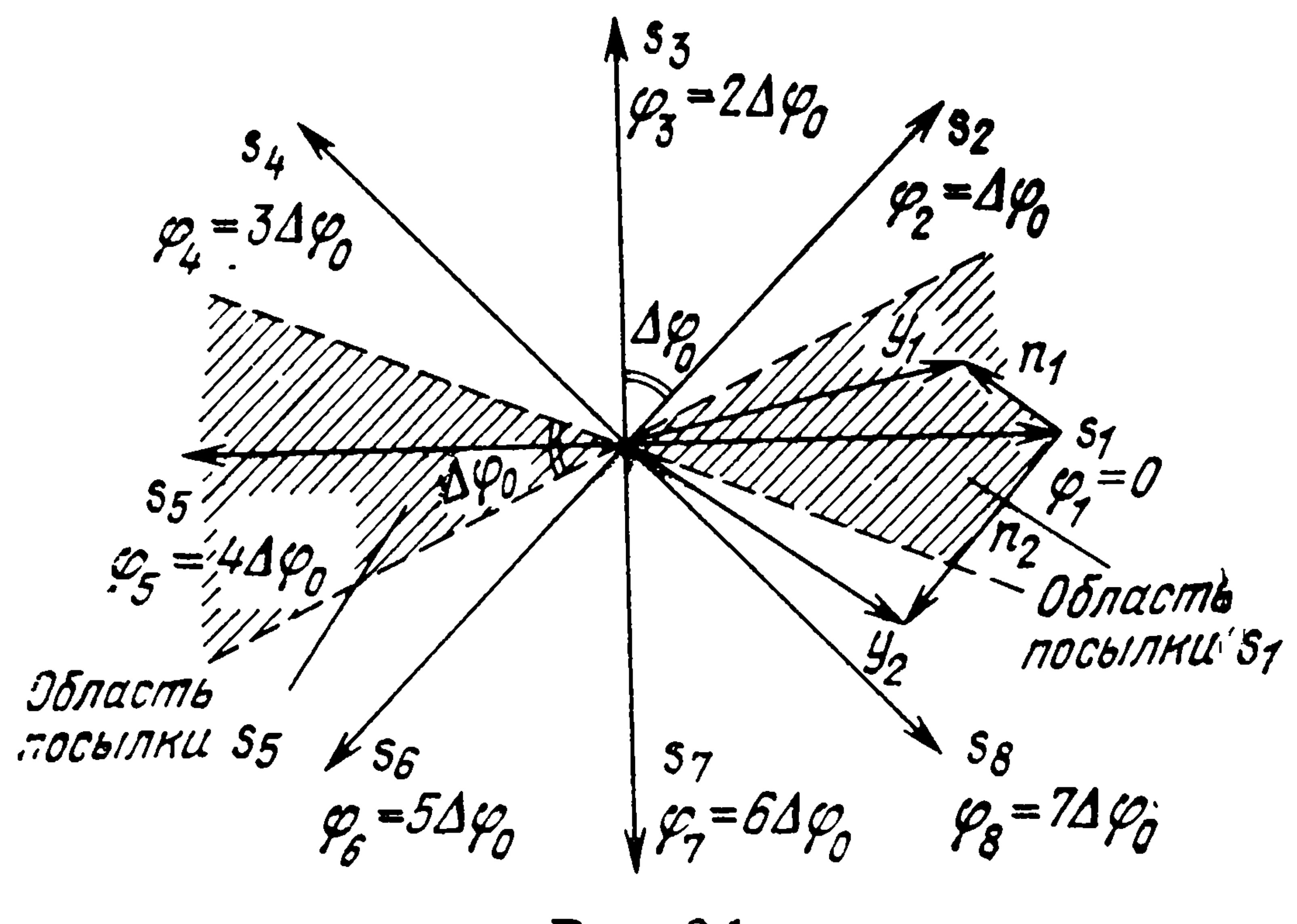


Рис. 6.1.

ные ошибки под действием шумовой помехи связаны с переходом переданной посылки в области соседних посылок. Например, переход посылки s_1 в область посылки s_2 или s_8.

Вероятность ошибочного приема символа, как и в предыдущем случае, приближенно определяется выражением (6.4). Для его конкретизации необходимо найти выражение для условной вероятности $p(s_{i-1}/s_i)$. Основываясь на результатах гл. 3 (см. (3.96)), нетрудно показать, что

$$p(s_{i-1}\,|\,s_i) = 0{,}5 - \Phi\left[\sqrt{\frac{S^2_0\tau_0}{2N_0}(1-\cos\Delta\varphi_0)}\right]. \qquad (6.10)$$

Приняв во внимание

$h^2_M = S^2_0\tau_0/2N_0$, $1-\cos\Delta\varphi_0 = 2\sin^2(\Delta\varphi_0/2) = 2\sin^2(\pi/M)$,

имеем

$$p(s_{i-1}\,|\,s_i) = 0{,}5 - \Phi\,[\sqrt{2}\,h_M\sin(\pi/M)]. \qquad (6.10а)$$

Величина

$$h^2_M \sin^2(\pi/M)$$

характеризует энергетические затраты на передачу $\log_2 M$ двоичных единиц информации. Следовательно, затраты на одну двоичную единицу составляют

$$h^2_0 = (h^2_M/\log_2 M)\sin^2(\pi/M).$$

Поступая как и в предыдущем случае, нетрудно найти из этого выражения необходимое отношение сигнал/шум, при котором обеспечивается требуемая достоверность передачи информации:

$$\left(\frac{P_c}{P_ш}\right)_{вх} \geqslant \frac{h^2_{отр}\log_2 M}{\sin^2(\pi/M)} = \frac{h^2_{отр}M^2\log_2 M}{\pi^2}\left[\frac{\pi/M}{\sin(\pi/M)}\right]^2 \quad (6.11)$$

Нетрудно показать, что при $M>4$ это выражение с хорошим приближением можно заменить следующим:

$$\left(\frac{P_c}{P_ш}\right)_{вх} \geqslant 0{,}1 h^2_{отр} M^2 \log_2 M. \quad (6.12)$$

Из сравнения (6.7) и (6.12) следует, что при одинаковых условиях многократная ФМн требует значительно меньшей мощности сигнала, чем многократная АМн. Кроме того, применение многократной АМн требует высокой линейности амплитудных характеристик выходных каскадов передатчика. Поэтому обычно в качестве многопозиционных сигналов, позволяющих эффективно использовать полосу частот, предпочитают использовать многократную ФМн. Однако в некоторых случаях возможно также применение и многопозиционной АМн, например при передаче по каналам с постоянными параметрами и низким уровнем шумов [1].

Необходимо отметить, что требования стабильности параметров канала и малого уровня помех существенны и для многократной ФМн. Эти требования резко возрастают при увеличении кратности манипуляции и часто становятся основной причиной, препятствующей такому увеличению на практике.

6.3. ПРИМЕРЫ ПРИМЕНЕНИЯ СИГНАЛОВ С МНОГОКРАТНОЙ ФАЗОВОЙ МАНИПУЛЯЦИЕЙ

Наибольшее применение сигналы с многократной ФМн находят в системах передачи данных по телефонным каналам. Прежде чем перейти к рассмотрению при-

меров таких систем, остановимся на принципах формирования сигналов с многократной ФМн и методах их приема.

Пусть имеется некоторая цифровая последовательность $x_1x_2x_2x_1x_2x_2x_1x_2x_1x_1$, образованная с помощью алфавита, состоящего из двух символов x_1 и x_2. Если эту последовательность разбить на блоки по два символа в каждом, то среди таких блоков будут встречаться четыре комбинации: x_1x_1, x_1x_2, x_2x_1, x_2x_2. Обозначим эти

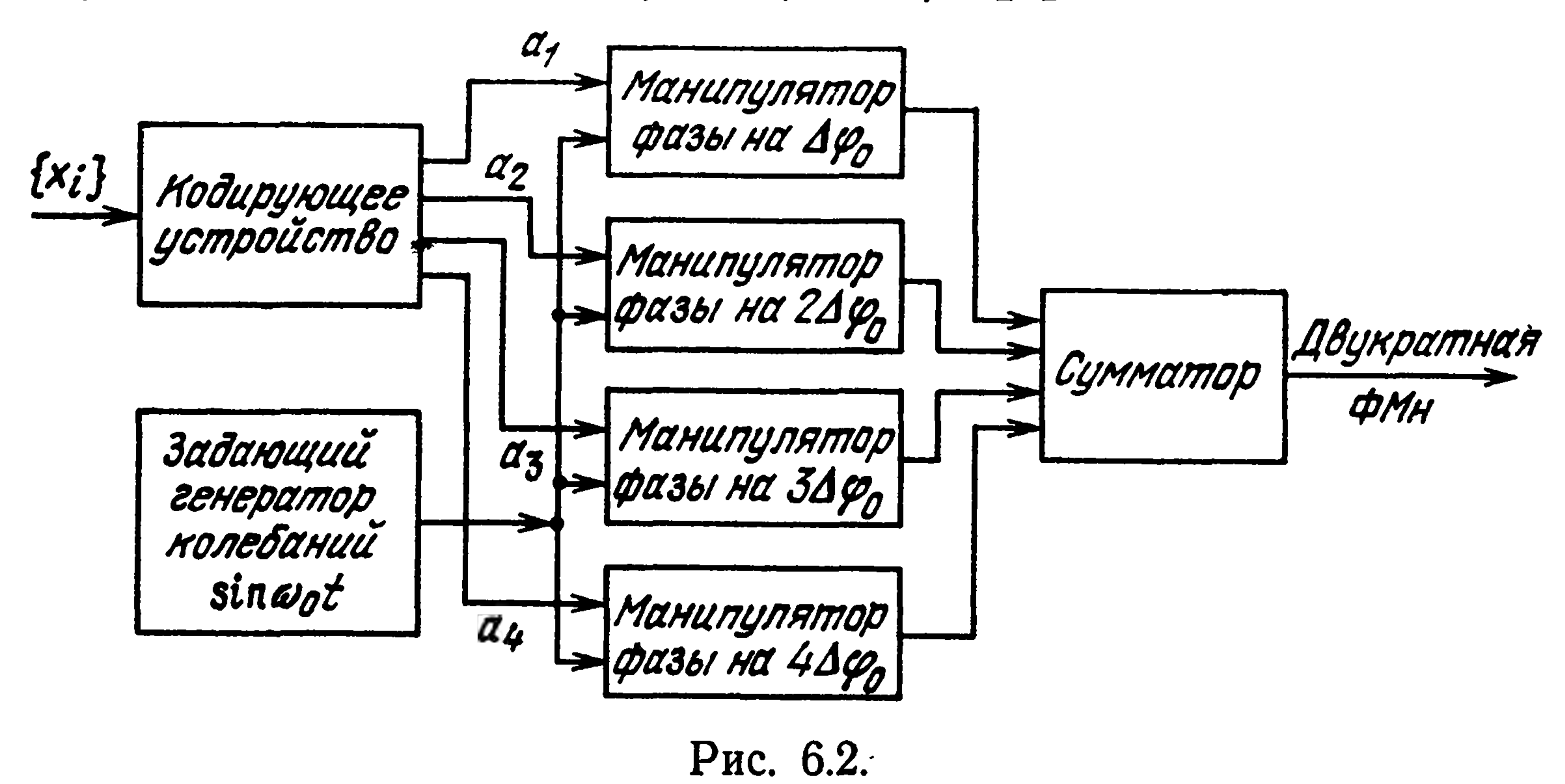


Рис. 6.2.

комбинации символами a_1, a_2, a_3, a_4. Тогда исходная последовательность примет вид $a_2a_3a_4a_2a_1$. Рассмотренная процедура соответствует перекодированию двоичной последовательности в четверичную.

Если для передачи исходной цифровой последовательности необходимо иметь две различающиеся посылки сигнала $x_1 \to s_1(t)$; $x_2 \to s_2(t)$, то для новой последовательности их должно быть уже четыре: $a_1 \to s_1(t)$, $a_2 \to s_2(t)$, $a_3 \to s_3(t)$; $a_4 \to s_4(t)$.

Нетрудно убедиться в том, что если разбивать исходную двоичную последовательность на блоки по k двоичных символов в каждом, то для передачи такой перекодированной последовательности потребуется $M=2^k$ различающихся посылок. Величина k определяет кратность ФМн сигнала, а следовательно, и минимальную величину фазового сдвига между посылками

$$\Delta\varphi_{\min}=\Delta\varphi_0=2\pi/M=2\pi/2^k.$$

На рис. 6.2 приведена упрощенная структурная схема передающей части системы с двукратной ФМн. Кодирующее устройство преобразует двоичную цифровую

последовательность в цифровую последовательность, образованную символами четверичного алфавита. Это достигается следующим образом. Каждому сочетанию входных символов $x_i x_j (i, j=1, 2)$ соответствует определенный символ a_m $(m=1, 2, 3, 4)$, который представляет собой либо импульс, либо отсутствие импульса. Например, сочетанию $x_1 x_1$ соответствует символ a_1 в виде импульса и символы a_2, a_3, a_4, равные нулю; сочетанию $x_2 x_1$ — символ a_3 в виде импульса и символы a_1, a_2 и a_4, равные нулю, и т. д.

В каждом манипуляторе начальная фаза исходного синусоидального колебания поворачивается с помощью фазовращателя на необходимый угол, и при появлении импульса с кодирующего устройства формируется посылка сигнала с данной фазой. В результате на выходе сумматора появится сигнал с двукратной ФМн, посылки которого однозначно соответствуют четверичным символам a_m. Передача каждого такого четверичного символа эквивалентна передаче двух двоичных символов.

Подобным же образом можно сформировать сигнал с двухкратной ФМн для одновременной передачи двоичной цифровой информации от двух независимых каналов, т. е. использовать этот сигнал для уплотнения двоичных каналов. В этом случае с помощью четверичного кода перекодируются сочетания двоичных символов, соответствующие разным каналам. Пусть, например, на интервале 0—τ в первом канале появился символ $x_{1i} (i=1, 2)$, а во втором — символ $x_{2j} (j=1, 2)$. Тогда такому сочетанию соответствует символ a_2, если $i=1$, а $j=2$; символ a_1, если $i=1$ и $j=1$, и т. д.

Рассмотрим теперь кратко некоторые особенности приема сигналов многократной ФМн.

Как известно, прием ФМн сигналов наталкивается на ряд трудностей, суть которых была рассмотрена в гл. 4. При многократной ФМн эти трудности усугубляются, поэтому передачу ведут, как правило, относительными методами фазовой манипуляции (ОФМн).

Для приема сигналов с многократной ОФМн можно использовать как корреляционные, так и автокорреляционные методы.

Рассмотрим особенности корреляционного приема на примере двукратной ОФМн (ДОФМн). Для реализации этого метода необходимо создать опорное напряжение из принимаемого сигнала, устранив манипуляцию фазы.

Особенностью является то, что устройства формирования опорного напряжения существенно усложняются по сравнению с устройствами для однократной ОФМн.

В частности, при использовании схемы Пистолькорса необходимо для устранения манипуляции фазы применять учетверение частоты и фазы принимаемого сигнала, а затем выполнять обратное преобразование — делить частоту получаемого колебания на четыре. В результате

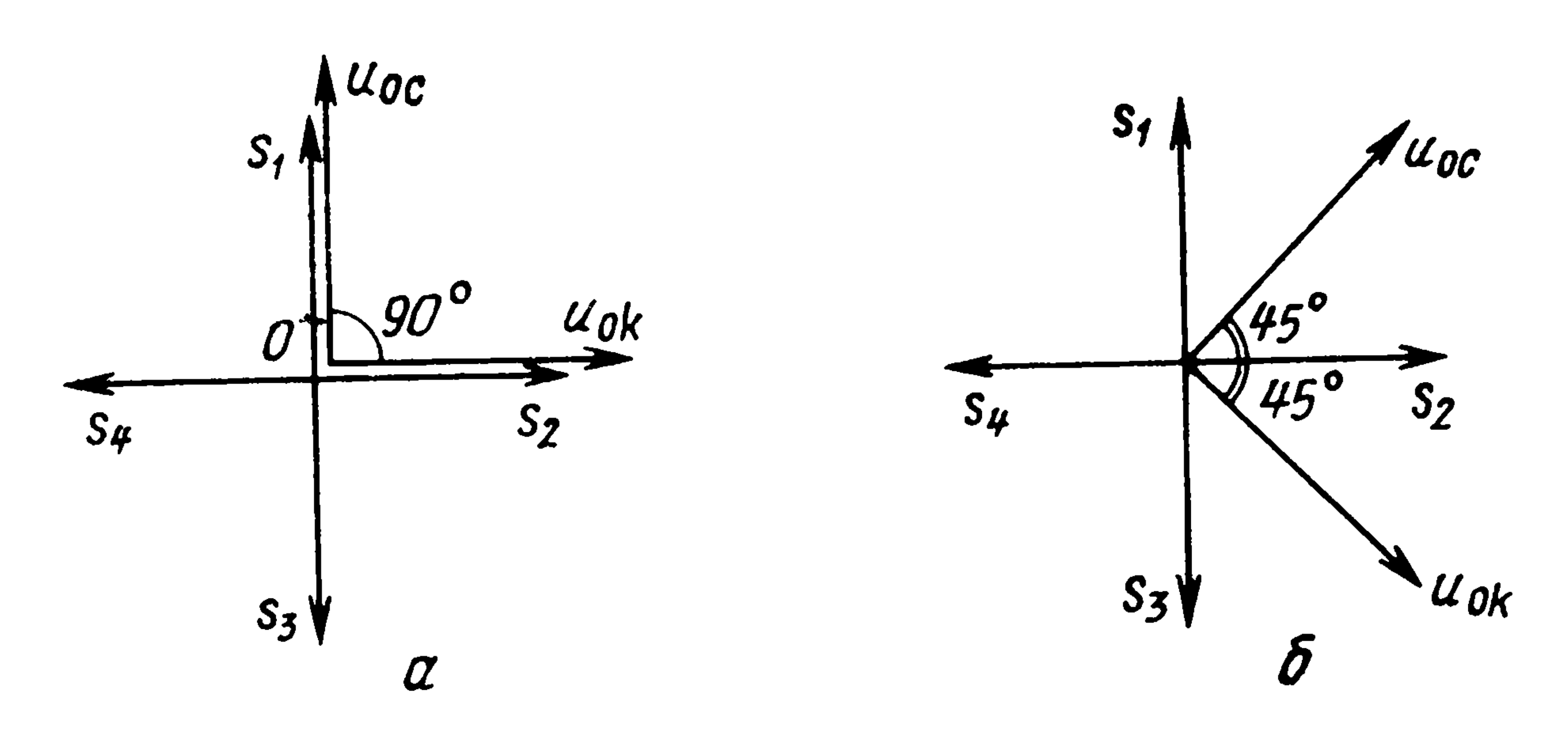


Рис. 6.3.

возникает неопределенность начальной фазы, которая может принять одно из четырех возможных значений: $\Delta\varphi_0$, $2\Delta\varphi_0$, $3\Delta\varphi_0$, $4\Delta\varphi_0$. Эта неопределенность начальной фазы опорного напряжения может привести либо к обратной работе, либо к неправильному приему символов (в зависимости от значения, которое примет начальная фаза). Такое положение характерно и для других известных способов формирования опорного напряжения.

Поскольку при ДОФМн фаза посылки может принять одно из четырех возможных значений, для демодуляции необходимо иметь два опорных напряжения, сдвинутых по фазе относительно друг друга на 90° и соответствующим образом сфазированных с приходящими посылками. Варианты такого фазирования приведены на рис. 6.3.

Упрощенная структурная схема демодулятора при использовании ДОФМн для фазового уплотнения двух двоичных каналов приведена на рис. 6.4. Фазовращатель (ФВ) обеспечивает необходимую начальную фазировку опорного напряжения относительно принимаемых посылок с учетом постоянных фазовых сдвигов, возникающих в среде и в избирательных цепях приемника. Для устранения влияния неопределенности начальной фазы опорного напряжения приходится применять более сложные логические устройства перекодирования по сравнению

с обычной (однократной) ОФМн. В таком устройстве обычно выполняется одновременно и разделение каналов.

Известен ряд вариантов реализации корреляционного метода приема многократных сигналов с ОФМн [2, 3].

Характерной особенностью автокорреляционного приема многократных сигналов с ОФМн является то,

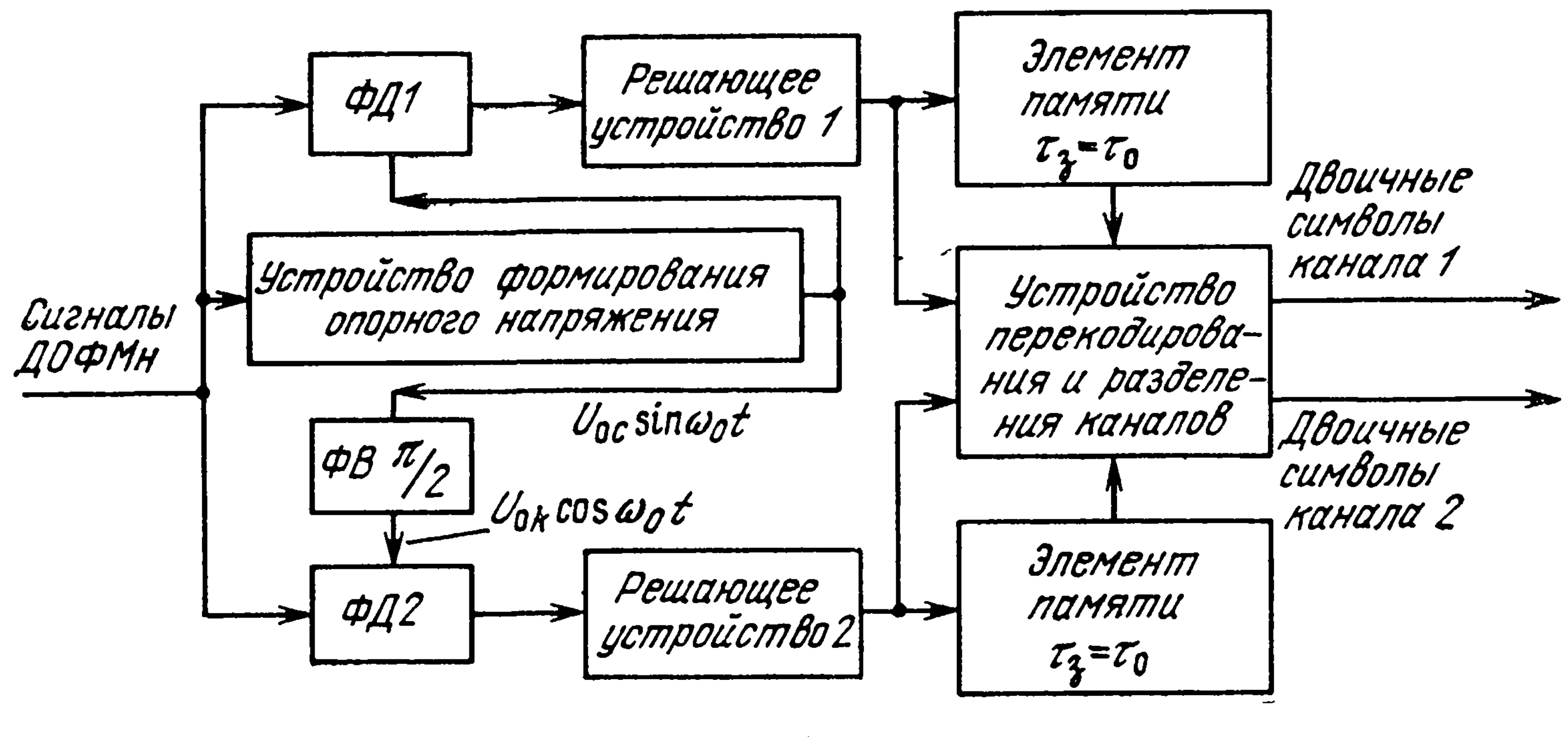


Рис. 6.4.

что для приема каждой посылки используется $M/2$ опорных напряжений, полученных из предыдущей задержанной посылки. Так, для приема ДОФМн нужно из каждой j-й посылки образовать два опорных напряжения, сдвинутых на 90°, которые используются при приеме $(j+1)$-й посылки. Как и при корреляционном приеме, схема демодулятора получается многоканальной (при ДОФМн — двухканальной). Однако демодулятор несколько проще, поскольку нет необходимости в устранении неопределенности фазы опорного напряжения *.

Весьма эффективный и сравнительно простой метод демодуляции, занимающий промежуточное положение между рассмотренными, описан в [4]. С вопросами схемных решений при создании модуляторов и демодуляторов систем с многократными ОФМн сигналами можно более подробно ознакомиться в [3, 5, 6].

Значительное увеличение необходимого отношения сигнал/шум на входе приемника, заметное усложнение схемы демодулятора, резкое повышение требований

* Напомним, что при автокорреляционном методе приема неопределенность фазы опорного напряжения принципиально невозможна (см. гл. 4).

к стабильности параметров сигнала и характеристик аппаратуры приемной части приводят к тому, что практическая реализация систем с сигналами высокой кратности наталкивается на большие трудности.

Наибольшее применение на практике в настоящее время находят системы с ДОФМн сигналами. В каналах с постоянными параметрами и низким уровнем помех применяют также сигналы с трехкратной ОФМн (ТОФМн). Более высокие кратности пока заметного применения не находят.

Классическим примером успешного практического применения сигналов с многократной ОФМн является система «Кинеплекс» [7]. Эта система была разработана в конце 50-х годов для высококачественной передачи цифровых данных и телеграфных сообщений по проводным или радиотелефонным стандартным каналам. Как известно, стандартный телефонный канал занимает полосу частот 300—3400 Гц. В этой полосе система «Кинеплекс» использует диапазон от 605 до 2696 Гц, который соответствует малым частотным и фазовым искажениям.

Передача цифровой информации осуществляется с помощью 20 поднесущих частот, расположенных в используемом диапазоне частот с разносом друг от друга на 110 Гц. Каждая поднесущая манипулируется по фазе методом ДОФМн, что обеспечивает передачу цифровой информации от двух разных источников на одной поднесущей. Таким образом, каждая поднесущая переносит информацию от двух разных каналов.

Разделение поднесущих частот при приеме осуществляется высокодобротными электромеханическими (кинематическими) фильтрами с добротностью порядка 1000. Одновременно эти же фильтры используются для запоминания фазы предшествующей посылки, необходимой для реализации автокорреляционного приема сигналов с ОФМн (см. гл. 4). Такой прием требует двух фильтров на каждой поднесущей. Коммутация фильтров осуществляется специальными синхронизирующими импульсами, действующими синфазно с началом и окончанием каждой информационной посылки. Для формирования синхронизирующих импульсов в системе используется специальный канал синхронизации на частоте 2915 Гц.

Принцип действия демодулятора ОФМн сигналов с кинематическими фильтрами был рассмотрен в гл. 4.

Поскольку в системе «Кинеплекс» каждая поднесущая манипулирована методом ДОФМн, схема демодулятора содержит не один, как при ОФМн, а два фазовых детектора, на выходе которых образуются цифровые последовательности отдельных каналов. Описание схемных особенностей системы «Кинеплекс» можно найти в [3, 4, 7].

Идеи, положенные в основу системы «Кинеплекс», оказались весьма плодотворными и получили дальнейшее развитие в ряде зарубежных и отечественных систем передачи цифровой информации по стандартным телефонным каналам. Известен ряд модификаций системы «Кинеплекс», которые обеспечивают скорость передачи 5400 бит/с и выше [4].

6.4. ОСОБЕННОСТИ СЛОЖНЫХ СИГНАЛОВ И ОСНОВНЫЕ ТРЕБОВАНИЯ К ТАКИМ СИГНАЛАМ

Теория информации указывает, что для построения систем передачи информации с малыми затратами энергии необходимо применить сложные сигналы (см. § 1.7).

Многочисленные исследования сложных сигналов позволили выяснить, что их особенности и свойства в существенной мере определяются величиной базы сигнала и проявляются наиболее ярко в тех случаях, когда эта база велика ($Б_с \gg 1$). Требование большой базы означает, что сложный сигнал занимает диапазон частот $\Delta f_э$, значительно превышающий минимально необходимый диапазон Δf_{min}, определяемый длительностью передаваемого информационного символа τ_0, т. е.

$$\Delta f_э \gg \Delta f_{min} = 1/\tau_0. \qquad (6.13)$$

Поэтому сложные сигналы, удовлетворяющие такому условию, часто называют широкополосными сигналами (ШПС) в отличие от узкополосных, для которых характерно соотношение $\Delta f_э \tau_0 \approx 1$. Широкополосность сигнала определяется не абсолютным значением ширины его спектра, а величиной базы. Величину базы сложного сигнала

$$Б_с = \Delta f_э / \Delta f_{min} = \Delta f_э \tau_0 \gg 1 \qquad (6.14)$$

можно рассматривать как некоторый параметр, характеризующий сознательно вводимую в сигнал избыточность. Именно она и определяет ряд специфических свойств сложных сигналов, которые открывают возможность для создания систем передачи информации, обладающих весьма ценными качествами.

Применение сложных сигналов позволяет реализовать информационные системы, с помощью которых успешно решают следующие задачи:

1) получение высокой достоверности передачи цифровой информации в условиях многолучевого распространения сигналов;

2) обеспечение высокой помехоустойчивости к организованным помехам и возможность криптостойкой и имитостойкой передачи особо важной информации*;

3) эффективное использование выделенного диапазона частот при одновременной работе в этом диапазоне многих систем, т. е. обеспечение электромагнитной совместимости различных систем, использующих один и тот же диапазон частот.

Хотя принципиальные возможности сложных сигналов следуют из общих положений теории связи, установленных 20—25 лет назад, применение таких сигналов для решения ряда задач началось позже (примерно 10—15 лет назад) сначала в радиолокации, а затем и в связи. В настоящее время исследованию и применению сложных сигналов уделятся большое внимание. Этому способствуют большие достижения в области микроэлектроники, вычислительной техники и т. п.

Цель рассмотрения этого и последующих параграфов состоит в том, чтобы дать основные представления о свойствах сложных сигналов, методах получения таких сигналов и методах их приема, а также привести примеры некоторых важных применений сложных сигналов.

В литературе используется много разнообразных терминов и названий сложных сигналов: широкополосные, псевдослучайные, шумоподобные, составные, многомерные. Каждый из этих терминов отображает какую-то из особенностей сложных сигналов, связанных с их свой-

---

* Термины «криптостойкость» и «имитостойкость» происходят от латинских слов kryptos — тайный, скрытый и imitatio — повторение, подделка. Они характеризуют скрытность передачи и трудность ее расшифровки, а также трудность создания помех, похожих на передаваемые сигналы, для дезориентации получателя информации.

ствами или методами их получения. В дальнейшем будем пользоваться термином «сложный сигнал» как наиболее общим.

При выборе необходимой величины базы сложного сигнала нужно различать два случая: 1) $\Delta f_э = \text{const}$, 2) $\tau_0 = \text{const}$.

В первом случае задана ширина спектра сложного сигнала. Такое положение возникает, если имеются ограничения на ширину используемого диапазона частот. В этом случае нужное значение базы можно получить, только увеличивая длительность информационного символа τ_0 (а следовательно, и длительность посылки сложного сигнала): $Б_с = \Delta f_э \tau_0 \equiv \tau_0$. При этом с увеличением базы (увеличением τ_0) техническая скорость передачи $R_т = 1/\tau_0$ будет падать. Следовательно, применение сложных сигналов в данном случае связано с уменьшением информативности системы и может быть оправдано в таких каналах, где информативность не имеет первостепенного значения и в первую очередь необходимо обеспечить высокую помехоустойчивость (каналы синхронизации, командные линии и т. п.).

Во втором случае задана длительность информационного символа τ_0 (а следовательно, и длительность соответствующей ему посылки сложного сигнала). При таком условии требуемую величину базы можно обеспечить только соответствующим выбором ширины спектра сложного сигнала. Техническая скорость передачи в этом случае остается постоянной и при соответствующем значении длительности символа τ_0 может быть весьма высокой. Применение сложных сигналов в данном случае может оказаться целесообразным в каналах, где отсутствуют ограничения на ширину полосы частот, занимаемой спектром сложного сигнала, но имеются случайные изменения параметров (например, в тропосферных каналах).

Применение сложных сигналов как в первом, так и во втором случае связано с большими удельными затратами полосы (см. § 1.7). Однако такие затраты в ряде случаев целесообразны, так как они позволяют получить системы с очень ценными свойствами.

Чтобы сложные сигналы могли эффективно использоваться в различных информационных системах, необходимо выполнить ряд требований, предъявляемых к таким сигналам. Рассмотрим эти требования.

Допустим, что в системе передачи цифровой информации используется n различных посылок сложных сигналов $s_1(t)$, $s_2(t)$, ..., $s_j(t)$, ..., $s_n(t)$. Каждой из посылок сигнала на входе приемника соответствует колебание, представляющее собой сумму этой посылки, искаженной в общем случае из-за неидеальных условий распространения в среде, и флюктуационного шума, т. е. $y_j(t)=s_{xj}(t)+n(t)$. Анализируя поступившее на вход колебание $y_j(t)$, приемник должен «решить», какая же посылка содержится в принятом колебании. Это задача распознавания n возможных сигналов или, иными словами, задача принятия одной гипотезы из n возможных.

Из теории статистических решений известно (а на частном примере различения двух сигналов, подробно рассмотренном в гл. 3, мы убедились в этом), что все статистически оптимальные процедуры различения n дискретных сигналов при наличии шума сводятся к сравнению n корреляционных интегралов:

$$I_k=\int_0^{\tau_0} y_j(t)\, s_k(t)\, dt, \qquad (6.15)$$

вычисляемых для всех значений $k=1, 2, \ldots, j, \ldots, n$. В (6.15) $y_j(t)$ — принятое колебание, образованное одной из посылок сигнала и шумом; $s_k(t)$ — образец сигнала, хранящийся в приемнике.

Если колебание $y_j(t)$ образовано посылкой $s_k(t)$ и шумом, то при правильно выбранных сигналах с большой вероятностью должно выполняться условие

$$\int_0^{\tau_0} y_k(t)\, s_k(t)\, dt \gg \int_0^{\tau_0} y_j(t)\, s_k(t)\, dt, \quad j \neq k. \qquad (6.16)$$

Выполнение этого условия означает, что набор сложных сигналов обладает хорошими корреляционными свойствами, т. е.

$$\int_0^{\tau_0} s_k(t)\, s_k(t)\, dt \gg \int_0^{\tau_0} s_j(t)\, s_k(t)\, dt, \quad j \neq k. \qquad (6.17)$$

Условие (6.17) показывает, что значение автокорреляционной функции сигнала при нулевом сдвиге ($\tau=0$) должно существенно превышать значение взаимно-кор-

реляционных функций *. Иными словами, автокорреляционная функция сложного сигнала должна содержать один значительный максимум (пик), а взаимно-корреляционная функция в любой момент времени — иметь существенно меньшую величину.

С большой вероятностью подобными свойствами обладают корреляционные и взаимно-корреляционные функции реализаций нормального шума конечной длительности [8, 9]. Поэтому структура сложных сигналов должна быть такой, чтобы их корреляционные свойства приближались к аналогичным свойствам отрезков шума. Из-за такого сходства сложные сигналы часто называют псевдошумными или шумоподобными.

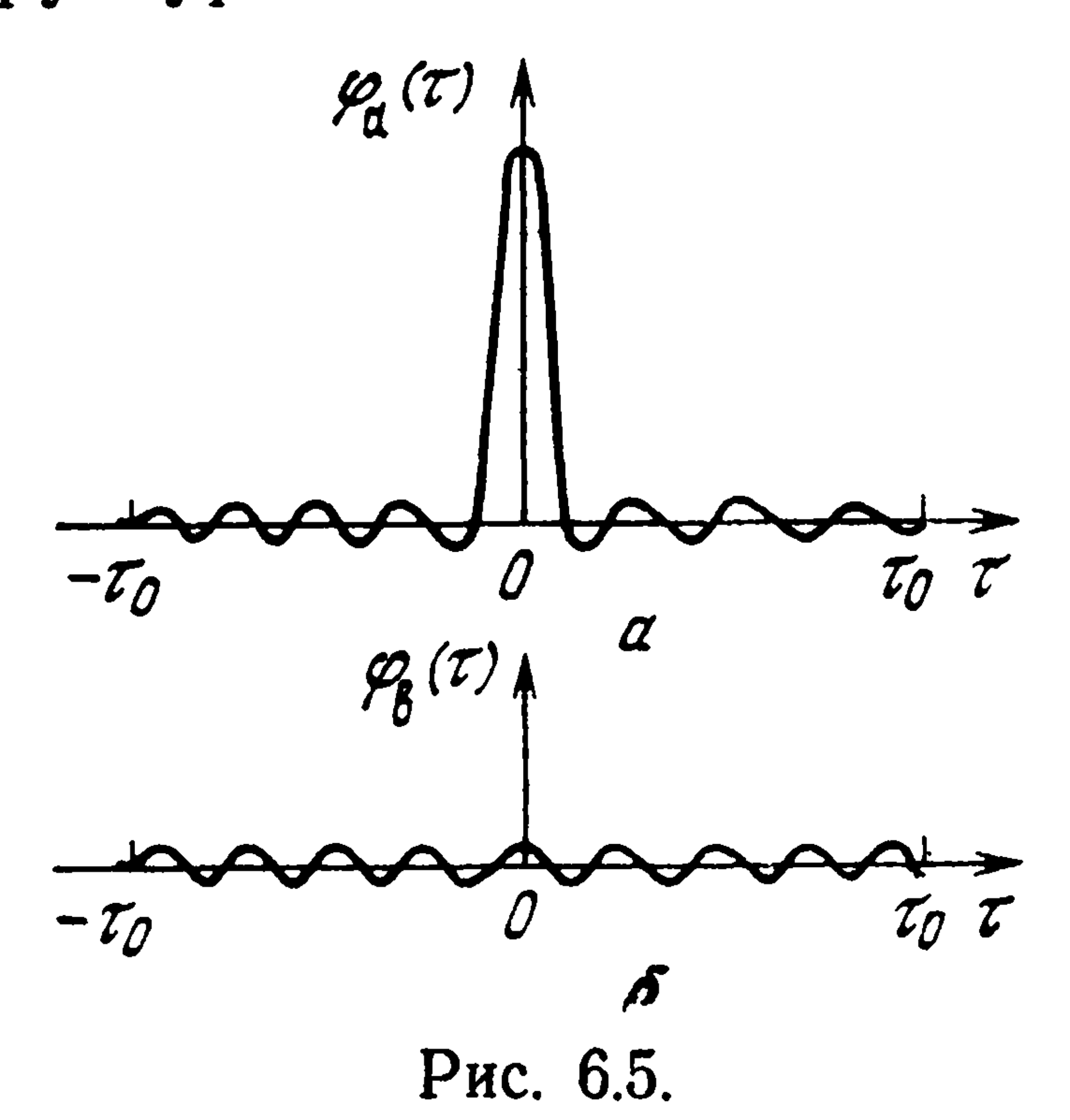


Рис. 6.5.

Итак, первым требованием к сложным сигналам является получение «хороших» корреляционных свойств совокупности посылок такого сигнала. В соответствии с этим требованием автокорреляционная функция сигнала должна иметь один значительный основной пик и минимальные остатки (боковые лепестки). Примерный вид желательной автокорреляционной и взаимно-корреляционной функции сложного сигнала длительностью τ_0 показан на рис. 6.5,*а* и *б* соответственно.

С учетом изложенного возникает естественный вопрос — нельзя ли в качестве необходимых посылок слож-

---

* Напомним, что автокорреляционной функцией сигнала $s_k(t)$ конечной длительности τ_0 называется выражение

$$\varphi_а(\tau) = \frac{1}{\tau_0}\int_0^{\tau_0} s_k(t)\, s_k(t-\tau)\, dt,$$

а взаимно-корреляционной функцией сигналов $s_j(t)$ и $s_k(t)$ выражение

$$\varphi_в(\tau) = \frac{1}{\tau_0}\int_0^{\tau_0} s_j(t)\, s_k(t-\tau)\, dt.$$

При $\tau = 0$ имеем $\varphi_а(0) \gg \varphi_в(0)$, что эквивалентно условию (6.17).

ного сигнала использовать набор независимых шумовых реализаций конечной длительности?

Обладая требуемыми корреляционными свойствами, шумовые реализации вместе с тем имеют крупный недостаток, связанный с их большим пикфактором. Как известно, пикфактором колебания $k_{п}$ называют отношение максимальной величины этого колебания x_{max} к эффективному значению $x_{эф}$. Применительно к сигналам можно записать

$$k_{п} = \frac{S_{max}}{S_{эф}} \text{ или } k^2{}_{п} = P_{пик}/P_{с}, \tag{6.18}$$

где $P_{пик}$ — пиковая мощность; $P_{с}$ — средняя мощность сигнала.

Нетрудно видеть, что $k_{п} \geqslant 1$. В реальных условиях необходимо считаться с ограниченностью пиковой мощности сигнала. В таких условиях сигналы с пикфактором, приближающимся к единице, обеспечивают хорошее использование мощности передатчика (высокий к. п. д. передатчика), а сигналы с большим пикфактором — плохое. Таким образом, шумовые реализации с энергетической точки зрения применять нецелесообразно.

Итак, вторым требованием к сложным сигналам является хорошее использование мощности передатчика.

Одним из важных условий успешного использования систем со сложными сигналами является необходимость оптимального (или близкого к нему подоптимального) приема информации. Сложность реализации такого приема до сих пор одна из основных трудностей, препятствующих более широкому внедрению указанных систем в практику. Как было установлено в гл. 3, для реализации оптимального приема необходимо располагать «образцами» используемых сигналов. Поэтому сложные сигналы должны допускать возможность воспроизведения их «образцов» на приемной стороне линии связи.

Помимо перечисленных, важным условием приема сложных сигналов является их четкая синхронизация.

Итак, сложные сигналы должны: 1) иметь большую базу; 2) обладать «хорошими» корреляционными свойствами; 3) иметь сравнительно небольшой пикфактор, т. е. обеспечивать хорошую «энергетику» излучаемого сигнала; 4) обеспечивать возможность воспроизведения в приемнике «образцов» используемых сигналов; 5) обеспечивать возможность четкой и надежной синхронизации в приемном устройстве.

6.5. МЕТОДЫ ПОЛУЧЕНИЯ СЛОЖНЫХ СИГНАЛОВ

В настоящее время известен ряд методов получения сложных сигналов [8—10]. Наибольший интерес для цифровых систем передачи информации представляет получение сложных сигналов с помощью псевдослучайных бинарных последовательностей (*М*-последовательностей) и с помощью частотно-временного кодирования. Достоинство этих методов в том, что они могут быть реализованы в основном с применением типовых элементов и устройств цифровой техники.

6.5.1. Получение сложных сигналов с помощью псевдослучайных бинарных последовательностей

Псевдослучайной бинарной последовательностью длительностью τ_0 называют последовательность, сформированную по определенным правилам из дискретных элементов 0 и 1 так, чтобы ее корреляционные свойства были близки к соответствующим свойствам шумовой реализации такой же длительности. Если обозначить длительность элемента последовательности через $\Delta\tau_0$, то число таких элементов в последовательности с длительностью τ_0 равно

$$M=\tau_0/\Delta\tau_0=Б_с\gg 1. \quad (6.19)$$

Такие последовательности в литературе часто называют *М*-последовательностями, а величину *М* — длиной последовательности.

Наиболее известны и хорошо исследованы *М*-последовательности, формируемые (генерируемые) двоичным регистром сдвига с логической обратной связью. Такой регистр представляет собой m последовательно соединенных триггерных ячеек, управляемых периодической последовательностью тактовых импульсов, вырабатываемых тактовым генератором, и импульсами, формируемыми в устройстве логической обратной связи.

Упрощенная структурная схема регистра показана на рис. 6.6 *. Рассмотрим работу этой схемы. Каждое плечо триггера (*Т*) может принимать два состояния — закрытое

* Ключи K_1 и K_2 в схеме рис. 6.6 введены условно и нужны лишь для пояснения ее работы.

и открытое (которые условно можно обозначать символами 0 и 1 или наоборот). Состоянием триггера будем называть состояние одного из его плеч (например, левого). Обозначим начальные состояния триггеров регистра через $\gamma_{01}, \gamma_{02}, \ldots, \gamma_{0j}, \ldots, \gamma_{0m}$. Ключи *K1* и *K2* при этом считаем разомкнутыми. Пусть в момент t_0 ключ *K1* замыкается и на все триггерные ячейки начинают поступать импульсы от генератора тактовых импульсов (ГТИ). Под действием этих импульсов триггеры изменяют свое состояние на противоположное или сохраняют прежнее состояние в зависимости от того, какое начальное состояние имел данный триггер и триггер, находящийся перед ним. В результате каждый триггер последовательно принимает все состояния предшествующих ему триггеров. Если снимать напряжение с последнего триггера, то оно последовательно принимает состояния $\gamma_{0m}, \gamma_{0(m-1)}, \ldots, \gamma_{0j}, \ldots, \gamma_{01}$, т. е. представляет последовательность, состоящую из m элементов, соответствующих символам 0 и 1. После m тактовых импульсов напряжение на выходе последней триггерной ячейки не меняется и она принимает состояние γ_{01}.

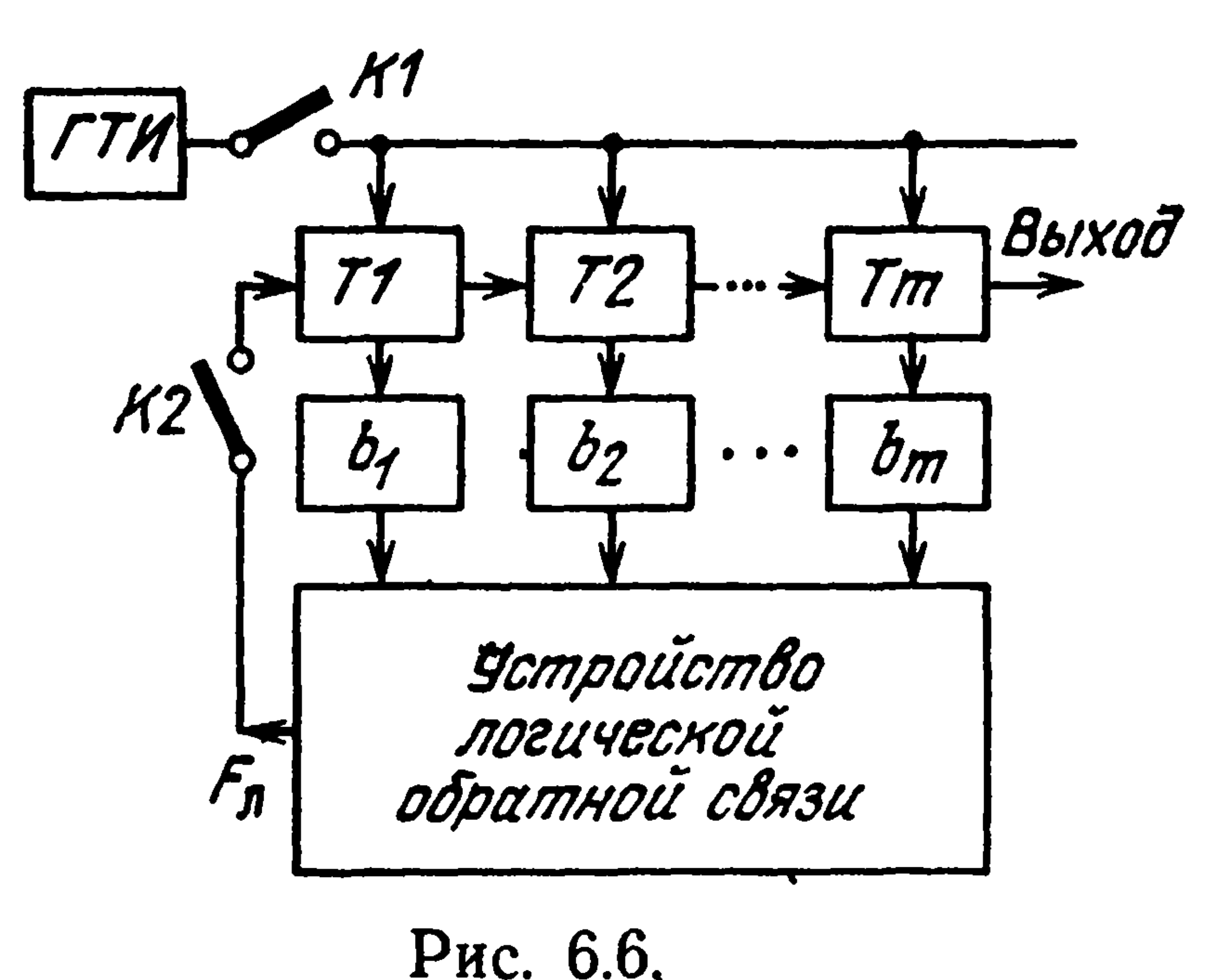


Рис. 6.6.

Таким образом, если регистр не имеет обратной связи (ключ *K2* разомкнут), то под действием тактовых импульсов на его выходе образуется только одна кодовая комбинация из m элементов, соответствующих начальным состояниям триггерных ячеек. В данном случае регистр представляет собой устройство памяти, в котором «записана» некоторая кодовая комбинация, которая может быть выведена (считана) в нужное время.

Чтобы регистр формировал дискретную последовательность непрерывно, т. е. работал как генератор такой последовательности, необходимо ввести обратную связь — замкнуть ключ *K2*. В этом случае в устройстве логической обратной связи формируется функция

$$F_л(\gamma_1, \gamma_2, \ldots, \gamma_j, \ldots, \gamma_m) = \sum_{j=1}^{m} b_j\gamma_j \pmod 2, \qquad (6.20)$$

которая называется логической функцией обратной связи. В (6.20) коэффициенты b_j могут принимать значения 1 или 0 в зависимости от того, какие ячейки триггеров участвуют в формировании функции $F_л$, определяющей обратную связь. Значения γ_j определяются фактическими состояниями триггерных ячеек в процессе работы.

Получение логической функции сводится к выбору совокупности коэффициентов b_j, определяющих, какие ячейки триггеров должны быть связаны с устройством логической обратной связи. Коэффициенты b_j выбираются в соответствии с определенным правилом, которое позволяет получить последовательность максимальной длины для регистра с заданным числом ячеек. Это правило может быть установлено на основе хорошо разработанной теории рекуррентных последовательностей [9, 11, 12]. Эти вопросы выходят за рамки данной книги, и мы на них останавливаться не будем. Укажем лишь, что имеются весьма подробные таблицы, которые содержат правила получения логической функции обратной связи $F_л$ для последовательностей различной длины [12].

Сформированная в соответствии с логической функцией дискретная последовательность символов 0 и 1 подается на вход первой триггерной ячейки регистра.

В результате действия тактовых импульсов и импульсов обратной связи регистр окажется в режиме непрерывной смены состояний, т. е. будет генерировать дискретную последовательность элементов. Общее число различных состояний, которые последовательно примет регистр сдвига, определяется числом триггерных ячеек m и тем, что каждая из них может находиться в одном из двух возможных состояний. Очевидно, что число различных состояний регистра (исключая нулевое состояние) равно

$$M = 2^m - 1. \qquad (6.21)$$

Следовательно, после каждых M разных состояний регистр, находящийся в режиме генерирования, начинает повторять эти состояния. Иными словами, регистр генерирует дискретную последовательность элементов, структура которой периодически повторяется, через каждые M элементов. Величина M поэтому называется периодом последовательности.

Регистр с обратной связью, образованной суммированием по «модулю два» состояний некоторых заранее

выбранных триггерных ячеек, называется линейным регистром сдвига или генератором псевдослучайной последовательности импульсов. Генерируемую последовательность можно снимать не только с выхода последнего триггера, но и с любого другого триггера. При этом последовательность имеет ту же структуру и отличается только сдвигом по времени.

Рассмотрим теперь работу линейного регистра сдвига, когда необходимо получить псевдослучайную последовательность импульсов, однозначно соответствующую исходной двоичной информационной последовательности символов.

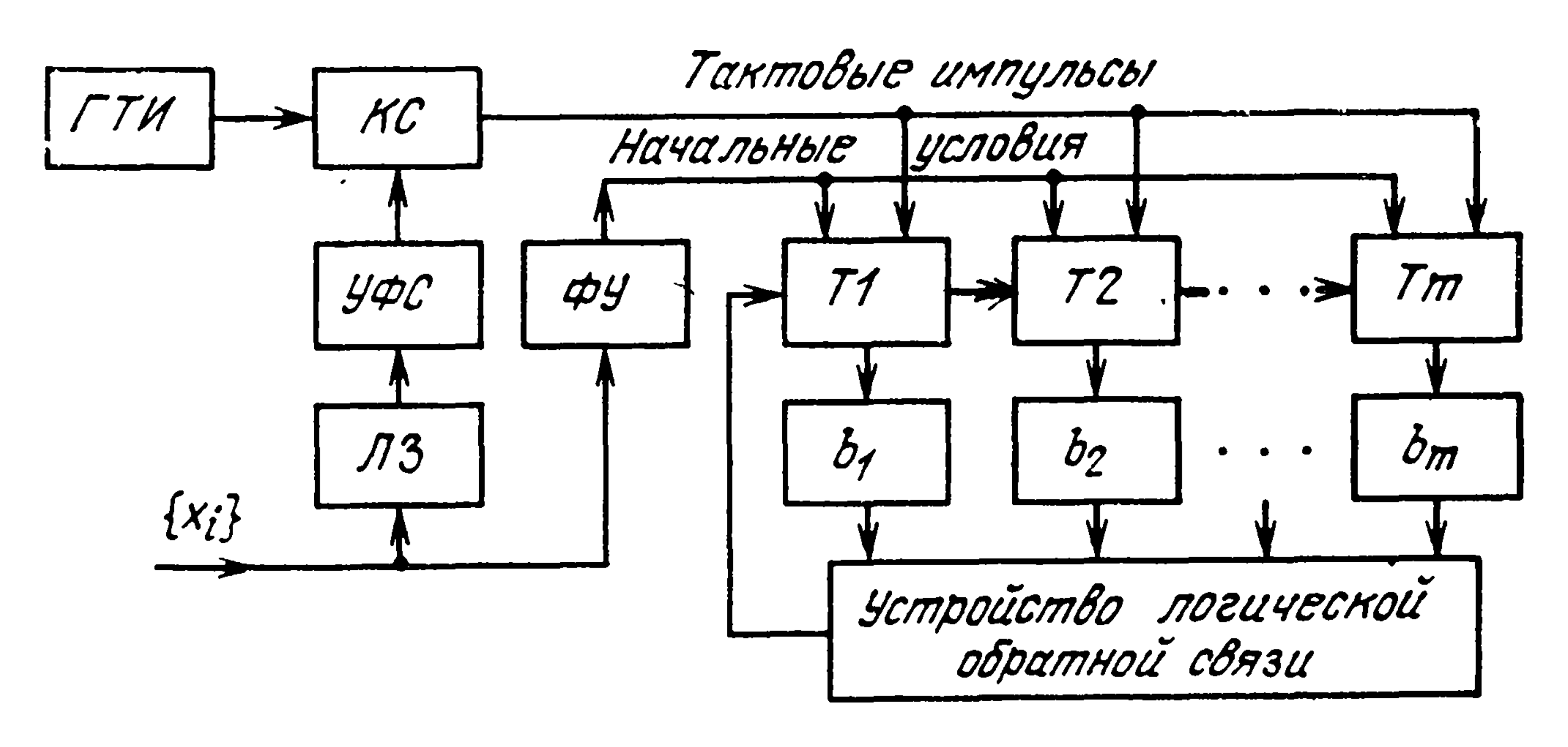


Рис. 6.7.

Выше указывалось, что перед началом работы регистра необходимо задать начальные условия в его ячейках, т. е. ввести в регистр некоторую начальную комбинацию символов. Эта комбинация определяет начало последовательности и ее внутреннюю структуру в пределах периода. Если перед началом каждого периода *М*-последовательности изменять начальные условия, то структура внутри каждого периода *М*-последовательности также будет изменяться. При передаче двоичной цифровой информации начальные условия должны изменяться в соответствии с исходной информационной последовательностью двоичных символов.

Один из возможных вариантов ввода исходной информационной последовательности в регистр сдвига приведен на рис. 6.7. Двоичная информационная последовательность символов $\{x_i\}$, каждый из которых имеет длительность τ_0, поступает на формирующее устройство (ФУ), вырабатывающее последовательность коротких

импульсов ($\tau_ф \ll \tau_0$), которые задают начальные условия во всех триггерных ячейках регистра. Во время задания начальных условий, соответствующих исходному информационному символу, ключевая схема (КС) закрыта и тактовые импульсы на регистр не поступают. Через некоторое время, определяемое линией задержки (ЛЗ) и необходимое для записи начальных условий, с выхода устройства формирования строба (УФС) на ключевую схему подается строб, открывающий эту схему для тактовых импульсов, и регистр начинает вырабатывать последовательность импульсов. В момент окончания строба ключевая схема закрывается и генерирование последовательности прекращается (регистр «останавливается»). После этого вводятся начальные условия, соответствующие следующему информационному символу, и схема начинает работать так же, как описано выше.

Проведенное рассмотрение показывает, что линейный регистр сдвига преобразует информационную последовательность простых двоичных символов в последовательность сложных двоичных символов. При этом каждому простому информационному символу 0 или 1 однозначно соответствует сложный символ 0 или 1, представляющий собой определенную комбинацию из M двоичных элементов. Если правило ввода начальных условий изменить (например, поменять некоторые плечи триггеров, на которые задаются начальные условия; не подавать начальные условия на некоторые триггеры и т. п.), то исходным символам 0 и 1 станут соответствовать сложные символы 0 и 1 с другой внутренней структурой. Нетрудно видеть, что в зависимости от задаваемых начальных условий одну и ту же исходную информационную последовательность простых двоичных символов регистр сдвига будет преобразовывать в последовательность с различными сложными двоичными символами.

Так как в зависимости от начальных условий регистр сдвига может принимать M различных состояний, то, очевидно, каждому простому символу 0 или 1 можно поставить в соответствие один из M сложных символов 0 или 1. Эта особенность работы регистра сдвига открывает ряд возможностей для решения различных прикладных задач.

Если полученную последовательность сложных информационных символов подать на модулятор и выполнить манипуляцию одного из параметров несущего ко-

лебания, то в результате образуется сложный сигнал, свойства и структура которого полностью определяются свойствами поданой последовательности и видом выбранной манипуляции.

Наибольшее распространение в настоящее время получили сложные сигналы, образованные манипуляцией начальной фазы несущей. Находят применение и сложные сигналы с частотной манипуляцией.

Манипулятор сложного сигнала с бинарной ФМн представляет собой быстродействующий коммутатор, осуществляющий переброс фазы несущего колебания в соответствии с последовательностью сложных символов на его входе. Сложный ФМн сигнал, соответствующий каждому сложному символу, состоит из M импульсов несущей частоты, начальные фазы которых равны нулю или 180°. Он может быть записан в виде

$$s_{сл}(t) = S_0 \sum_{k=0}^{M-1} f(t - k\Delta\tau_0) \exp[j(\omega_0 t + \pi\gamma_k)] \quad (6.22)$$

или

$$s_{сл}(t) = S_0 \sum_{k=0}^{M-1} (-1)^{\gamma_k} f(t - k\Delta\tau_0) \exp(j\omega_0 t) =$$

$$= S_0 \exp(j\omega_0 t) \sum_{k=0}^{M-1} (-1)^{\gamma_k} f(t - k\Delta\tau_0), \quad (6.23)$$

где $f(t)$ и $\Delta\tau_0$ огибающая и длительность элемента сложного сигнала; S_0 — амплитуда сложного сигнала; γ_k — случайная величина, принимающая значение 0 или 1 и однозначно соответствующая порядку следования двоичных элементов M-последовательности.

Обычно огибающую элемента считают прямоугольной:

$$f(t) = \begin{cases} 1, & 0 \leqslant t \leqslant \Delta\tau_0, \\ 0 & \text{для других } t. \end{cases}$$

Иногда в литературе прямоугольную огибающую обозначают знаком rect *.

В этом случае сигнал можно записать так:

$$s_{сл}(t) = S_0 \exp(j\omega_0 t) \sum_{k=0}^{M-1} (-1)^{\gamma_k} \operatorname{rect}(t - k\Delta\tau_0). \quad (6.24)$$

* От латинского слова rectangula — прямоугольная.

Выражение (6.24) показывает, что сложный ФМн сигнал является результатом перемножения немодулированной несущей на псевдослучайную M-последовательность (с учетом знака, определяемого множителем $(-1)^{\gamma_k}$). Величину

$$S(t) = S_0 \sum_{k=0}^{M-1} (-1)^{\gamma_k} \operatorname{rect}(t - k\Delta\tau_0) \qquad (6.25)$$

можно рассматривать как комплексную огибающую сложного сигнала. Спектр сигнала отличается от спектра этой огибающей только переносом на частоту ω_0. Поэтому при изучении спектральных и корреляционных

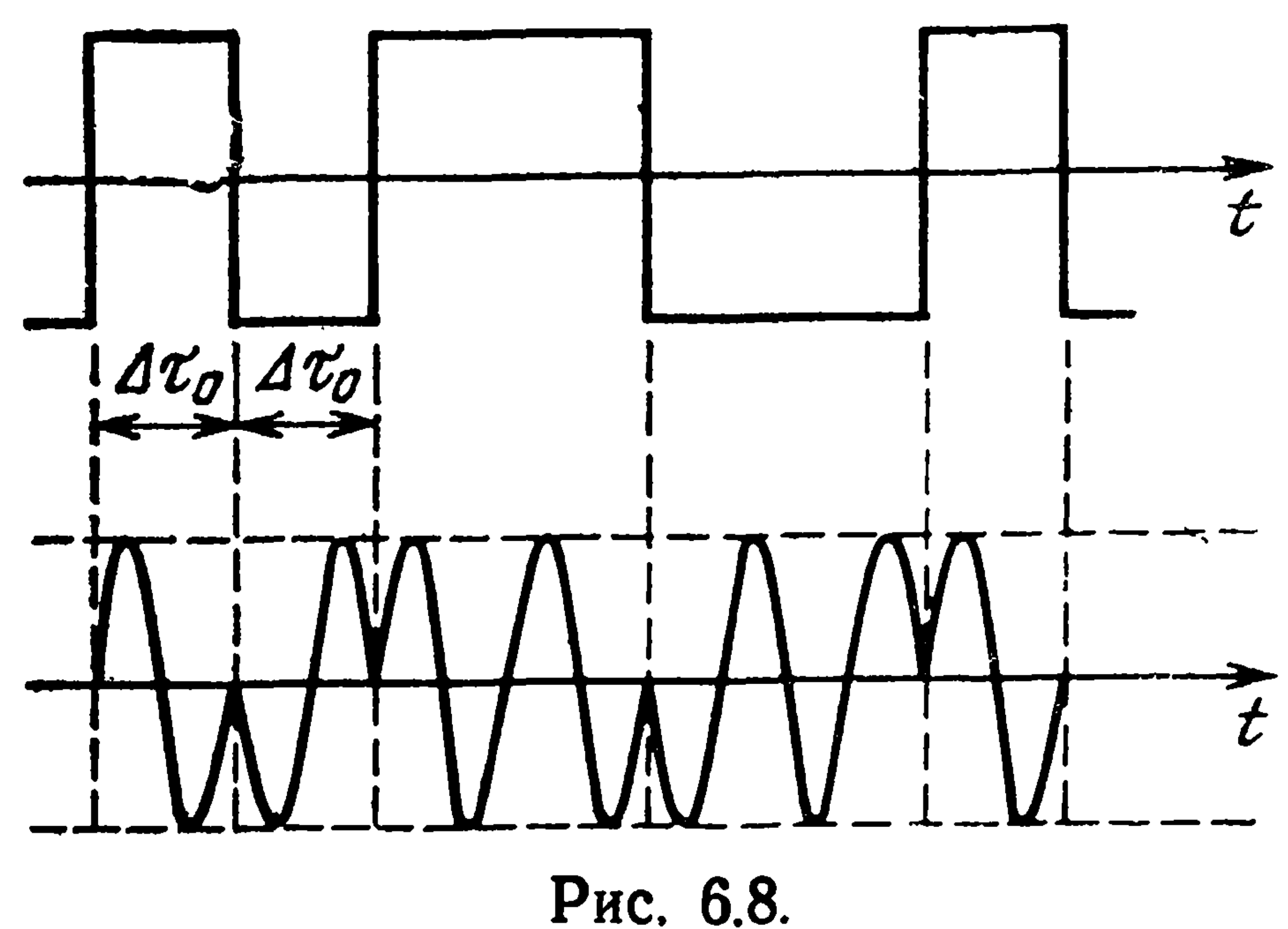


Рис. 6.8.

свойств сложного ФМн сигнала достаточно рассматривать только его комплексную огибающую (6.25). Она имеет вид знакопеременной последовательности вида $++-+-++-\ldots$ или $1,1-1,1-1,1,1,-1,\ldots$, однозначно определяющей порядок чередования начальных фаз (0 или π) элементов сложного сигнала.

Для иллюстрации на рис. 6.8 показана часть огибающей и соответствующие ей элементы сложного сигнала.

Таким образом, выбор сложного ФМн сигнала (или его синтез) сводится к выбору (синтезу) последовательности $\{\gamma_k\}_M$, обладающей требуемыми свойствами. Эту последовательность иногда называют кодом сложного сигнала.

Аналогичное рассмотрение можно провести и применительно к получению сложного сигнала с ЧМн. В этом случае каждый сложный символ манипулирует частоту несущего колебания, которая в зависимости от вида двоичного элемента сложного символа принимает значение f_1 или f_2.

Многочисленные исследования показали, что свойства M-последовательностей близки к свойствам случайных последовательностей. В пределах периода каждую M-последовательность можно рассматривать как одну из реализаций некоторой случайной последовательности. Чем больше длина периода M, тем ближе это сходство.

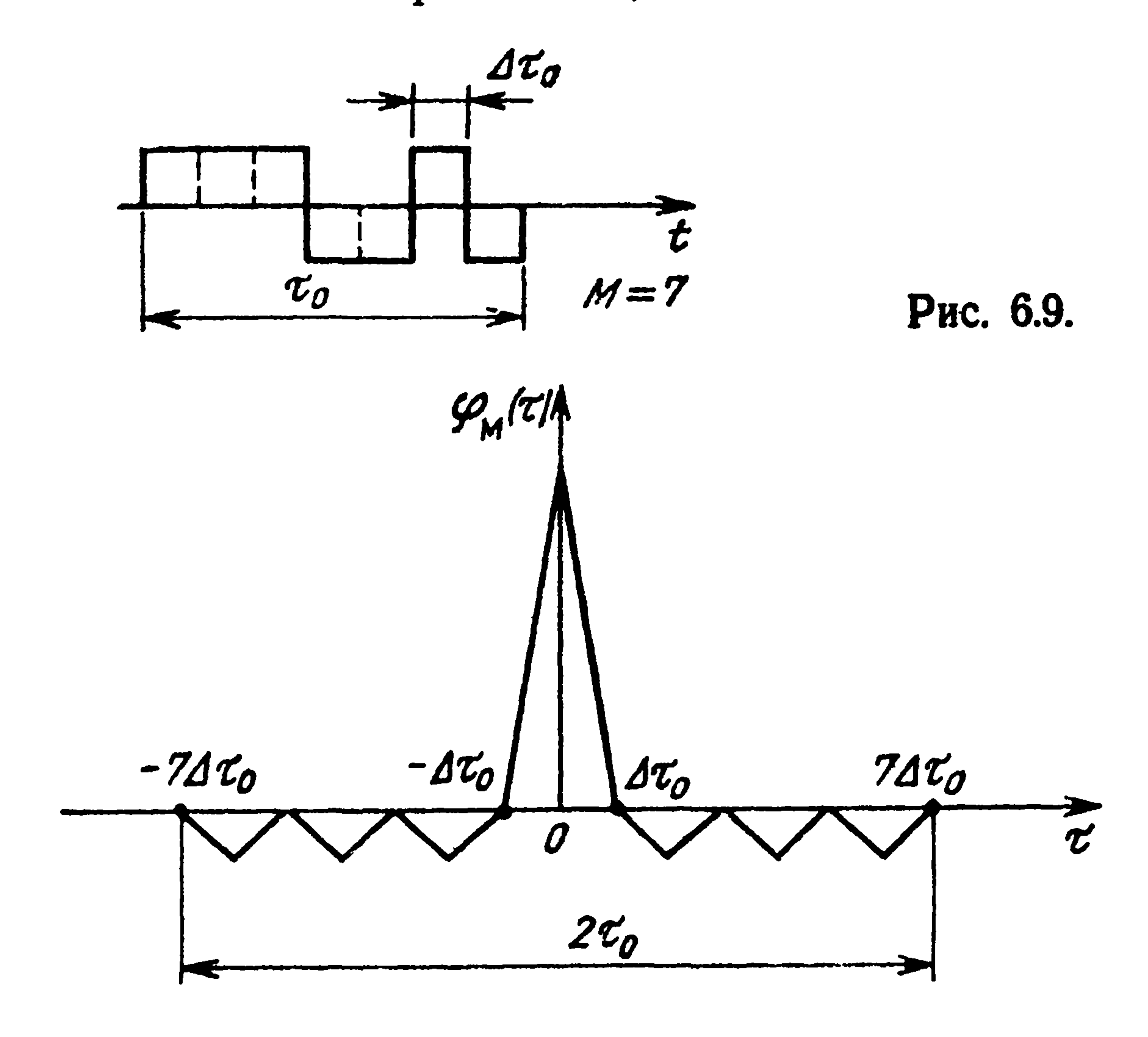


Рис. 6.9.

Поскольку, однако, природа M-последовательностей не случайна и они могут быть воспроизведены многократно (при известном алгоритме их получения), их называют псевдослучайными последовательностями.

Свойства M-последовательностей подробно рассмотрены в работах [9, 10, 12]. Мы рассмотрим кратко только корреляционные свойства M-последовательностей, используемых в качестве сложных информационных символов. Функцию автокорреляции $\varphi_M(\tau)$ M-последовательности, имеющей длительность τ_0, можно найти по известному выражению

$$\varphi_M(\tau) = \frac{1}{\tau_0}\int_0^{\tau_0} S(t)\,S(t+\tau)\,dt. \qquad (6.26)$$

Здесь $S(t)$ определяет аналитическое выражение M-последовательности как функции времени.

В частности, для ФМн сложного сигнала эта функция выражается соотношением (6.25). Нетрудно убедиться в том, что функция автокорреляции имеет вид лома-

ной линии, изменяющей свой наклон в точках $\tau=k\Delta\tau_0$. При этом в точке $\tau=0$ функция принимает максимальное значение; в точках $\tau=k\Delta\tau_0$, где k принимает нечетные значения ($k=\pm1$, ±3, ±5 и т. д.), функция равна нулю. Для всех промежуточных значений τ функция $\varphi_M(\tau)$ отлична от нуля.

На рис. 6.9 для иллюстрации приведена M-последовательность ($M=7$) и соответствующая ей функция автокорреляции.

Чтобы сложный сигнал имел «хорошие» корреляционные свойства, боковые лепестки (остатки) функции $\varphi_M(\tau)$ должны быть малы по сравнению с основным лепестком (пиком). Для анализа остатков достаточно вычислить значения функции $\varphi_M(\tau)$ только в дискретных точках, соответствующих моментам $\tau=k\Delta\tau_0$ для целых значений k ($k=0$, ±1, ±2 и т. д.), т. е. найти

$$\varphi_M(\tau)=\varphi_M(k\Delta\tau_0)=\frac{1}{M}\sum_{i=1}^{M-k}\gamma_k\gamma_{i+k}, \qquad (6.27)$$

где γ_k — значение элемента M-последовательности ($+1$ или -1).

Аналогичное рассмотрение можно провести и для функции взаимной корреляции различных M-последовательностей длительностью τ_0. Эта функция имеет вид ломаной линии, все пики которой имеют приблизительно одинаковую величину.

Основным показателем корреляционных свойств одиночных M-последовательностей является максимальный уровень остатков автокорреляционной функции. Исследования показывают, что наименьший максимальный уровень остатков определяется выражением [10]

$$|\varphi_M(\tau)|_{\min}=\varphi_M(0)/M \text{ при } |\tau|>\Delta\tau_0\,(M\leqslant 13). \qquad (6.28)$$

Последовательности, которые обеспечивают такие остатки, называются оптимальными. К классу таких последовательностей относятся последовательности, изученные впервые в 1953 г. английским ученым Р. Баркером. В литературе они получили название кодов Баркера. Многочисленные исследования показали, что коды Баркера существуют только для $M\leqslant13$. Поэтому их применение для формирования сложных сигналов ограничено тем, что основной пик получается довольно ши-

роким. Однако коды Баркера очень удобны для создания сигналов синхронизации в цифровых системах связи с простыми сигналами, в радиолокации и т. п. Для M-последовательностей с большим числом элементов ($M>13$) в настоящее время не известны сигналы с уровнем остатков меньших, чем

$$\varphi_M(\tau)=\varphi_M(0)/\sqrt{M} \text{ при } |\tau|>\Delta\tau_0 \; (M>13), \quad (6.29)$$

Коэффициент корреляции между любыми одиночными M-последовательностями удобно определять с помощью соотношения

$$\rho_M(k\Delta\tau_0)=\frac{M_1-M_2}{M_1+M_2}=\frac{M-2M_2}{M}=1-\frac{2M_2}{M}, \quad (6.30)$$

где M_1 — число позиций, в которых элементы последовательностей совпадают; M_2 — число позиций, в которых элементы последовательностей различаются; k указывает, на какое число элементов сдвинуты последовательности относительно друг друга.

Нетрудно убедиться в том, что коэффициент взаимной корреляции между сложными символами 0 и 1, вырабатываемыми линейным регистром сдвига, равен —1, т. е. принимает минимально возможное значение. Это означает, что при прочих равных условиях передача цифровой информации противофазными двоичными сложными сигналами обеспечивает максимальную их различимость, а следовательно, и помехоустойчивость.

Проведенное рассмотрение показывает, что M-последовательности просты в реализации и обеспечивают получение свойств, необходимых для формирования сложных сигналов с требуемыми характеристиками.

6.5.2. Получение сложных сигналов с помощью частотно-временного кодирования

Допустим, что имеется M различных простых сигналов одинаковой мощности, каждый из которых имеет длительность $\Delta\tau_0$. Назовем эти сигналы элементарными и запишем их в виде

$$\begin{aligned}
s_1(t)&=S_0\sin(\omega_1 t+\varphi_1),\\
&\dots\dots\dots\dots\\
s_j(t)&=S_0\sin\{[\omega_1+(j-1)\Delta\omega_0]t+\varphi_j\}, \quad (6.31)\\
&\dots\dots\dots\dots\\
s_M(t)&=S_0\sin\{[\omega_1+(M-1)\Delta\omega_0]t+\varphi_M\}.
\end{aligned}$$

Здесь $\Delta\omega_0 = 2\pi\Delta f_0$ — некоторый постоянный сдвиг частоты; $\varphi_1, \varphi_2, \ldots, \varphi_M$ — начальные фазы.

Образуем последовательность из M непрерывно следующих друг за другом элементарных сигналов. При этом условимся, что порядок следования элементарных сигналов может быть любым, но каждый из таких сигналов должен входить в последовательность только один раз. При таком правиле формирования последовательности ее длительность равна

$$\tau_0 = M\Delta\tau_0, \tag{6.32}$$

а ширина занимаемой полосы частот (ширина спектра последовательности)

$$\Delta f_э = (M-1)\Delta f_0 + 1/\Delta\tau_0. \tag{6.33}$$

Полученную последовательность можно рассматривать как сложный сигнал, составленный из элементарных сигналов. Поэтому такой сигнал иногда называют составным. Величина базы этого сигнала равна

$$Б_с = \Delta f_э\tau_0 = M[1 + (M-1)\Delta f_0\Delta\tau_0]. \tag{6.34}$$

Осуществляя перестановку элементарных сигналов, можно получать различные последовательности, состоящие из M элементов. Такие последовательности можно представить как комбинации, образованные из элементов матрицы, показанной на рис. 6.10. Строки этой матрицы соответствуют разным частотам элементарных сигналов, а столбцы — разным временным позициям этих сигналов. Сигналы в одной строке имеют одну и ту же частоту и разное временное положение, а сигналы в одном столбце — одно и то же временное положение и различные частоты. Каждый элемент матрицы представляет собой элементарный сигнал, соответствующий данной временной и частотной позициям.

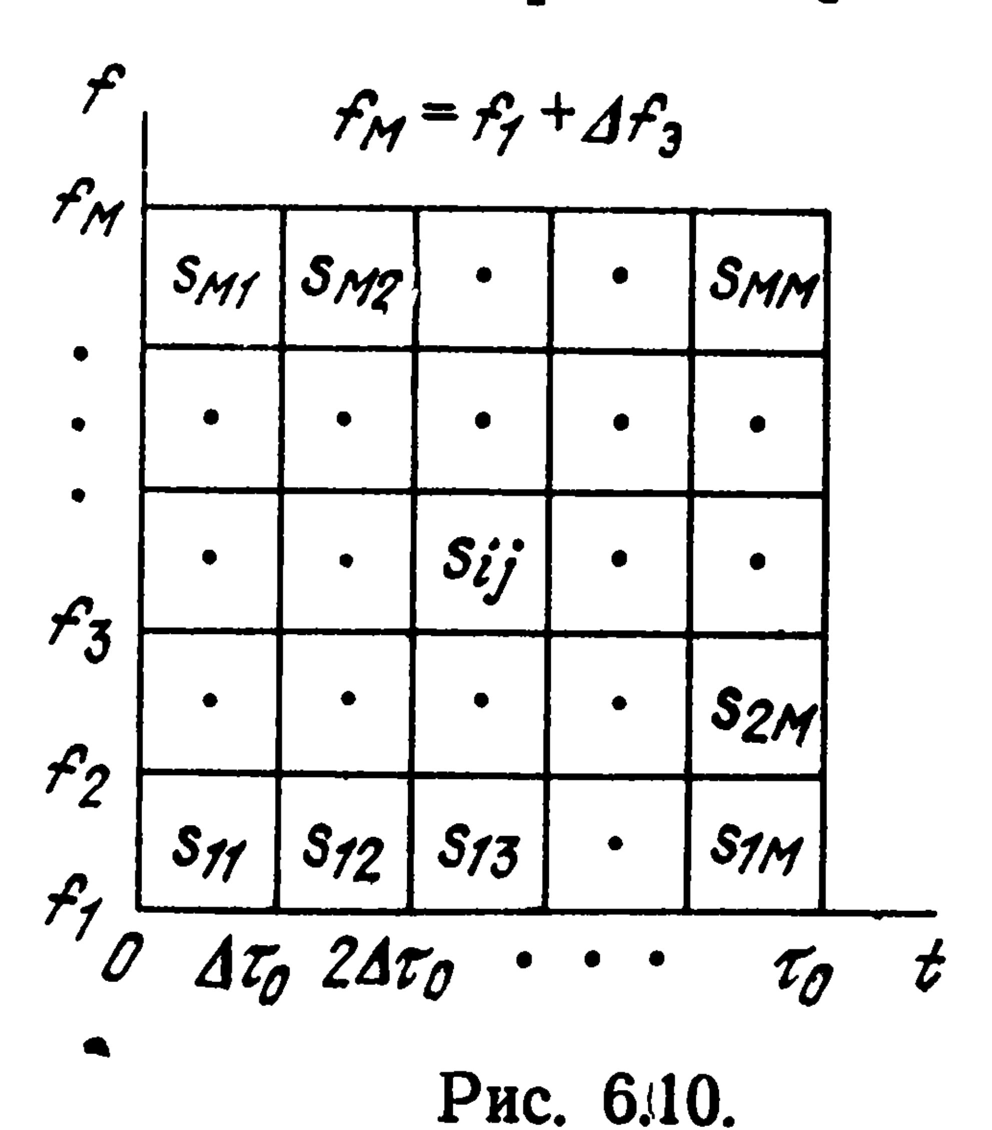


Рис. 6.10.

При формировании различных последовательностей на них, как правило, налагаются различные ограничения. Например, в каждой последовательности используются только те элементарные сигналы, которые относятся к разным столбцам и строкам, а использова-

ние более одного сигнала из каждой строки и столбца запрещается.

Рассмотренный принцип получения сложного (составного) сигнала называют частотно-временным кодированием, а используемую исходную совокупность элементов — частотно-временной матрицей.

Число различных перестановок элементов квадратной матрицы размером M с учетом указанных выше ограничений определяет количество различающихся друг от друга последовательностей длиной M. Это число равно

$$N=1\cdot 2\ \dots\ M=M! \qquad (6.35)$$

Отсюда следует, что частотно-временное кодирование позволяет получить значительное число различающихся сложных сигналов даже при небольших величинах M. Так, $N=120$ при $M=5$; $N=720$ при $M=6$ и т. д. Однако не все из полученных составных сигналов обладают достаточно хорошими авто- и взаимокорреляционными свойствами. Очевидно, что среди последовательностей будут такие, у которых элементарные сигналы на некоторых временных позициях совпадают. Это приводит к появлению заметных остатков во взаимокорреляционных функциях и ухудшает свойства таких последовательностей. Перебор различных вариантов и выбор последовательностей с нужными свойствами можно осуществить с помощью ЭВМ.

Основываясь на результатах, приведенных в работах [13, 14], можно показать, что при частотно-временном кодировании на основе квадратной матрицы размером M верхняя грань числа последовательностей с различающимися элементарными сигналами во всех временных позициях (ортогональные последовательности) определяется величиной

$$\sup N_{\text{opt}}=[2M-1]. \qquad (6.36)$$

Это соотношение показывает, что величина N_{opt} значительно меньше числа последовательностей, определяемых выражением (6.35). Поэтому практически приходится использовать последовательности, у которых некоторые временные позиции содержат одинаковые элементарные сигналы.

Генерирование сложных сигналов, формируемых на основе частотно-временного кодирования, сводится

к получению ступенчатых функций длительностью τ_0, принимающих различные дискретные значения на интервалах $\Delta\tau_0$, и использованию этих функций для манипуляции частоты несущего колебания. Примерный вид одной из таких функций показан на рис. 6.11. Различные ступенчатые функции можно рассматривать как реализации случайного процесса. Поэтому сложные сигналы будут обладать теми же свойствами, что и соответствующие им ступенчатые функции.

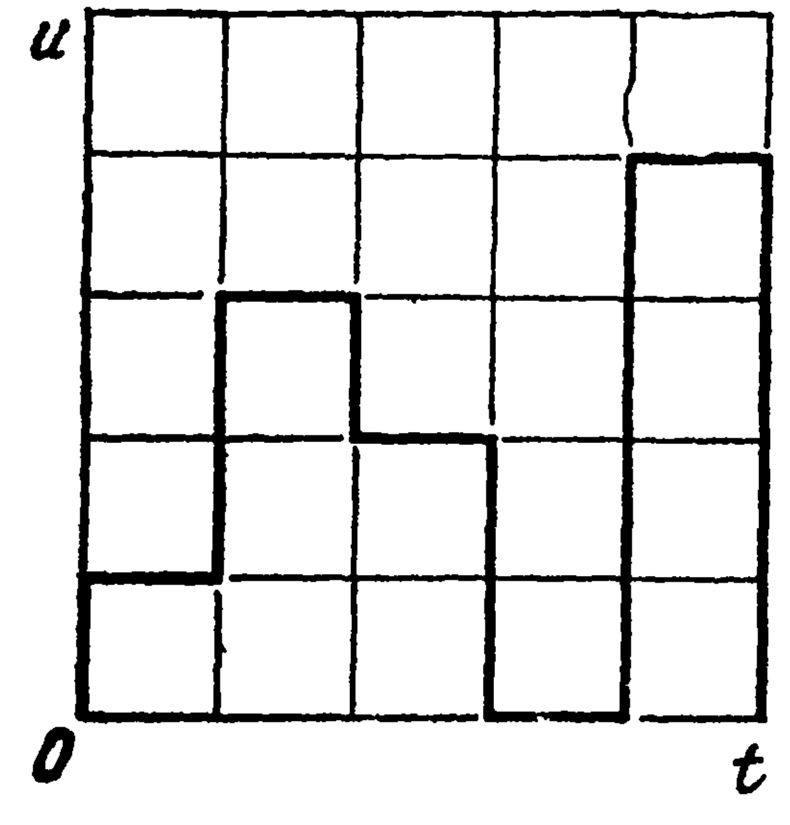


Рис. 6.11.

С некоторыми вопросами частотно-временного кодирования можно ознакомиться в работах [8, 9, 13]. Вопросы генерирования квантованных случайных процессов изложены в работах [15, 16].

6.6. МЕТОДЫ ПРИЕМА СЛОЖНЫХ СИГНАЛОВ

Как уже указывалось, одной из основных причин, затрудняющих применение сложных сигналов, являются трудности реализации приема таких сигналов. Чтобы положительные свойства сложных сигналов проявлялись наиболее полно, необходимо применять способы приема, близкие к оптимальным.

Рассмотрим сначала особенности оптимального приема сложных сигналов, а затем, опираясь на результаты такого рассмотрения, дадим краткую характеристику особенностей реальных методов приема. При этом для простоты ограничим рассмотрение случаем равновероятных двоичных сложных сигналов. Такие сигналы применяются наиболее широко, поскольку использование сигналов с более высокими основаниями связано с существенными усложнениями аппаратуры, которая уже и для двоичных сложных сигналов оказывается достаточно сложной.

В зависимости от уровня априорных сведений о параметрах принимаемого сигнала (начиная от сигнала, известного точно и кончая сигналом со случайными параметрами) возможен ряд методов оптимального приема сложных сигналов на фоне нормальной флюк-

туационной помехи типа белого шума [9, 10, 17]. Остановимся на двух наиболее важных методах.

6.6.1. Оптимальный когерентный прием

В соответствии с теорией оптимального приема равновероятных двоичных сигналов, известных точно, правило работы приемника определяется выражением

$$\int_0^{\tau_0} y(t)\, s_{1сл}(t)\, dt \gtrless \int_0^{\tau_0} y(t)\, s_{2сл}(t)\, dt \quad \begin{matrix} \hat{x}_i = x_1, \\ \hat{x}_i = x_2, \end{matrix} \tag{6.37}$$

где $s_{1сл}(t)$ и $s_{2сл}(t)$ — двоичные сложные сигналы, соответствующие передаваемыми символами x_1 и x_2.

Из рассмотрения правила (6.37) следует, что для оптимального приема двоичных сложных сигналов, известных точно, полностью справедливы все результаты, полученные в гл. 3. Особенностью является лишь то, что в качестве «образцов» на приемной стороне системы связи теперь необходимо располагать такими же сложными сигналами, которые используются для передачи.

Процедура оптимального приема может быть несколько упрощена, если отказаться от непосредственного приема сложного сигнала «в целом» и вести когерентный поэлементный прием сигнала с последующей обработкой полученных результатов.

Приняв во внимание, что посылки сложного сигнала можно представить в виде

$$s_{iсл}(t) = \sum_{k=1}^{M} s_{ik}(t - k\Delta\tau_0), \quad i = 1, 2, \tag{6.38}$$

где $s_{ik}(t)$ — элемент сложного сигнала, получим

$$\sum_{k=1}^{M} \int_0^{\tau_0} y(t)\, s_{1k}(t - k\Delta\tau_0)\, dt \gtrless \sum_{k=1}^{M} \int_0^{\tau_0} y(t)\, s_{2k}(t - k\Delta\tau_0)\, dt \quad \begin{matrix} \hat{x}_i = x_1, \\ \hat{x}_i = x_2. \end{matrix} \tag{6.39}$$

Это выражение показывает, что оптимальный прием сводится к когерентной обработке всех элементов посылки сигнала, суммированию полученных результатов и принятию решения о переданном символе в соответствии с результатом сравнения образованных сумм.

Для конкретизации выражения (6.39) необходимо задать вид используемого сигнала. При наиболее часто

применяемых противофазных сложных сигналах элементы $s_{1k}(t)$ и $s_{2k}(t)$ могут принимать только два возможных значения

$$s_{ik}(t)=\begin{cases} S_0 \sin \omega_0 t \\ S_0 \sin (\omega_0 t+\pi)\end{cases}, \quad 0<t<\Delta\tau_0. \qquad (6.40)$$

Тогда с учетом рассмотрения, проведенного в предыдущем параграфе, нетрудно убедиться в справедливости следующей записи:

$$s_{ik}(t-k\Delta\tau_0)=S_0(-1)^{\gamma_{ik}}\operatorname{rect}(t-k\Delta\tau_0)\sin\omega_0 t \qquad (6.41)$$

или

$$s_{ik}(t-k\Delta\tau_0)=\varphi_{ik}(t)S_0\sin\omega_0 t, \qquad (6.42)$$

где $\varphi_{ik}(t)$ — двоичная функция, принимающая значения $+1$ и -1 и определяющая последовательность смены фаз элементов сложного сигнала.

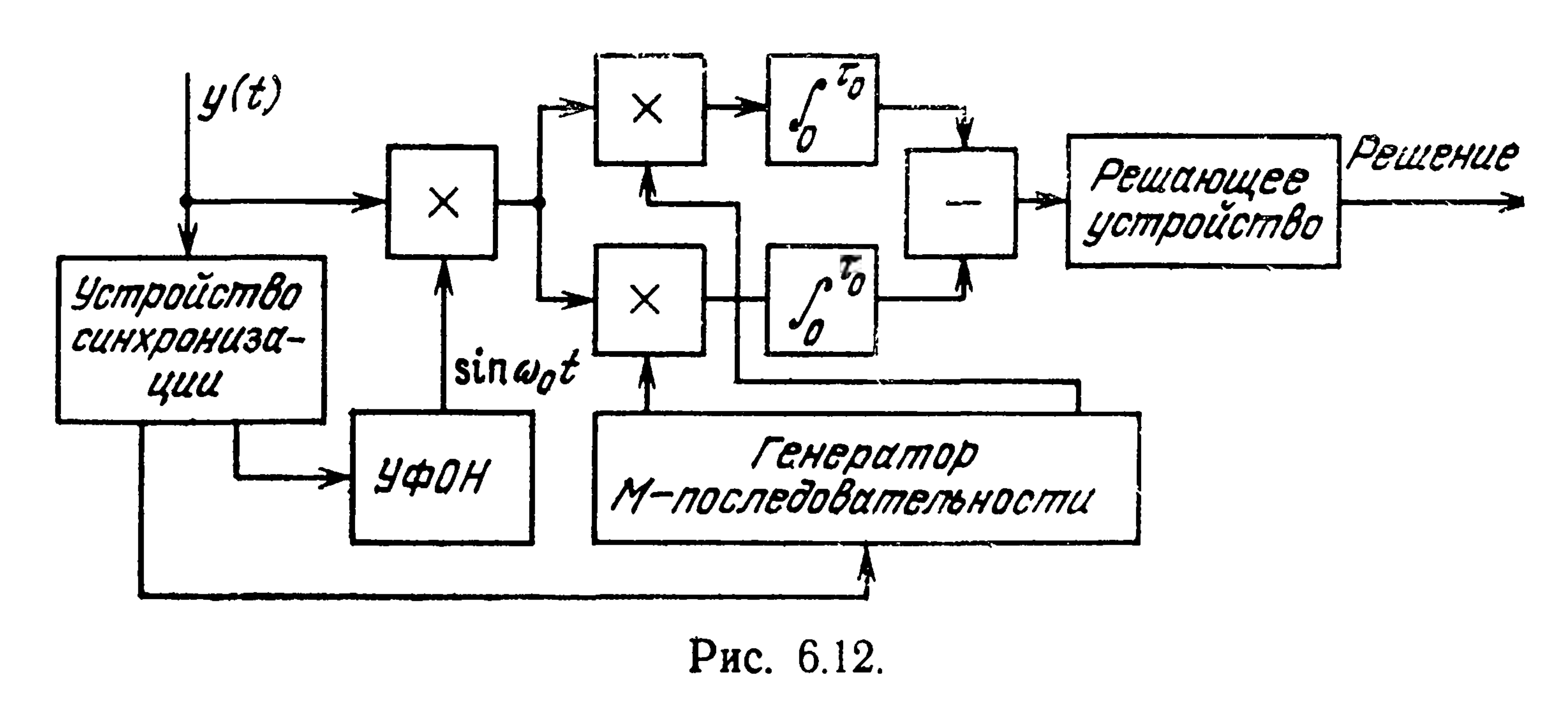


Рис. 6.12.

Приняв во внимание (6.42), выражение (6.39) для двоичного сложного сигнала с противофазной ФМн можно записать в виде

$$\sum_{k=1}^{M}\int_0^{\tau_0} y(t)\varphi_{1k}(t)\sin\omega_0 t dt \gtrless \sum_{k=1}^{M}\int_0^{\tau_0} y(t)\varphi_{2k}(t)\sin\omega_0 t dt \quad \begin{matrix}\hat{x}_i=x_1,\\ \hat{x}_i=x_2.\end{matrix} \qquad (6.43)$$

Это правило можно реализовать различными оптимальными схемами приема. Одна из них приведена на рис. 6.12.

Достоинством этой схемы является то, что она не требует «образцов» в виде сложных сигналов. Для приема достаточно иметь гармоническое опорное напряжение с нулевой фазой и псевдослучайную последова-

тельность, формируемую по такому же правилу, как и на передающей стороне системы. В одном канале приемника используется последовательность, снимаемая с одного плеча выходной ячейки регистра сдвига, а в другом канале — с другого плеча.

Хотя принципиально все варианты оптимального приема обеспечивают одинаковые результаты, их значимость с точки зрения практической реализации различна. Дело в том, что при практической реализации всегда имеется некоторое рассогласование между параметрами принимаемого сигнала и параметрами приемника. С увеличением базы сигнала значительно повышаются требования к точности изготовления и стабильности узлов и элементов приемника. Невыполнение этих требований может привести к заметному снижению помехоустойчивости по сравнению с оптимальным приемом. В этой связи особое значение приобретает выбор такого варианта оптимального приема, реализация которого обеспечивала бы значительное ослабление влияния дестабилизирующих факторов на помехоустойчивость. Исследованию указанной проблемы в настоящее время уделяется большое внимание. Одно из перспективных направлений связано с изучением методов дискретной обработки сложных сигналов. Прием в соответствии со схемой рис. 6.12 также относится к этой группе методов.

Если случайные рассогласования между параметрами сигнала и приемника таковы, что априорными сведениями о фазе сигнала воспользоваться нельзя (сигнал имеет случайную фазу), лучшие результаты обеспечивает оптимальный некогерентный прием.

6.6.2. Оптимальный некогерентный прием

Допустим, что цифровая информация передается двоичными сложными сигналами, определяемыми выражением (6.38). Если из-за различных факторов начальная фаза элементов таких сигналов приобретает случайный характер, то при противофазных сигналах каждый элемент сигнала можно записать в виде

$$s_{ik}(t)=s_{ik}(t-k\Delta\tau_0)=S_0\varphi_{ik}(t)\sin[\omega_0 t+\varphi_x(t)]. \quad (6.44)$$

Это выражение отличается от (6.42) только тем, что фаза элемента сигнала имеет дополнительный случайный сдвиг $\varphi_x(t)$.

Полагая случайные изменения фазы медленными, т. е. удовлетворяющими условию

$$\varphi_x(t) \approx \varphi_x = \text{const}, \qquad 0 \leqslant t \leqslant \tau_0,$$

и основываясь на методике, изложенной в § 5.3, можно показать, что правило работы оптимального приемника в этом случае определяется выражением [17, 27]

$$V_1 \gtrless V_2 \quad \begin{matrix} \hat{x}_i = x_1, \\ \hat{x}_i = x_2, \end{matrix} \tag{6.45}$$

где

$$V_i = \left\{ \left[\sum_{k=1}^{M} \int_0^{\tau_0} y(t)\, \varphi_{ik}(t) \sin \omega_0 t\, dt \right]^2 + \left[\sum_{k=1}^{M} \int_0^{\tau_0} y(t)\, \varphi_{ik}(t) \cos \omega_0 t\, dt \right]^2 \right\}^{1/2}, \quad i = 1, 2. \tag{6.46}$$

Так же, как и в предыдущем случае, возможен ряд вариантов схемы оптимального приемника, соответствующего правилу (6.45). Один из таких вариантов приведен на рис. 6.13.

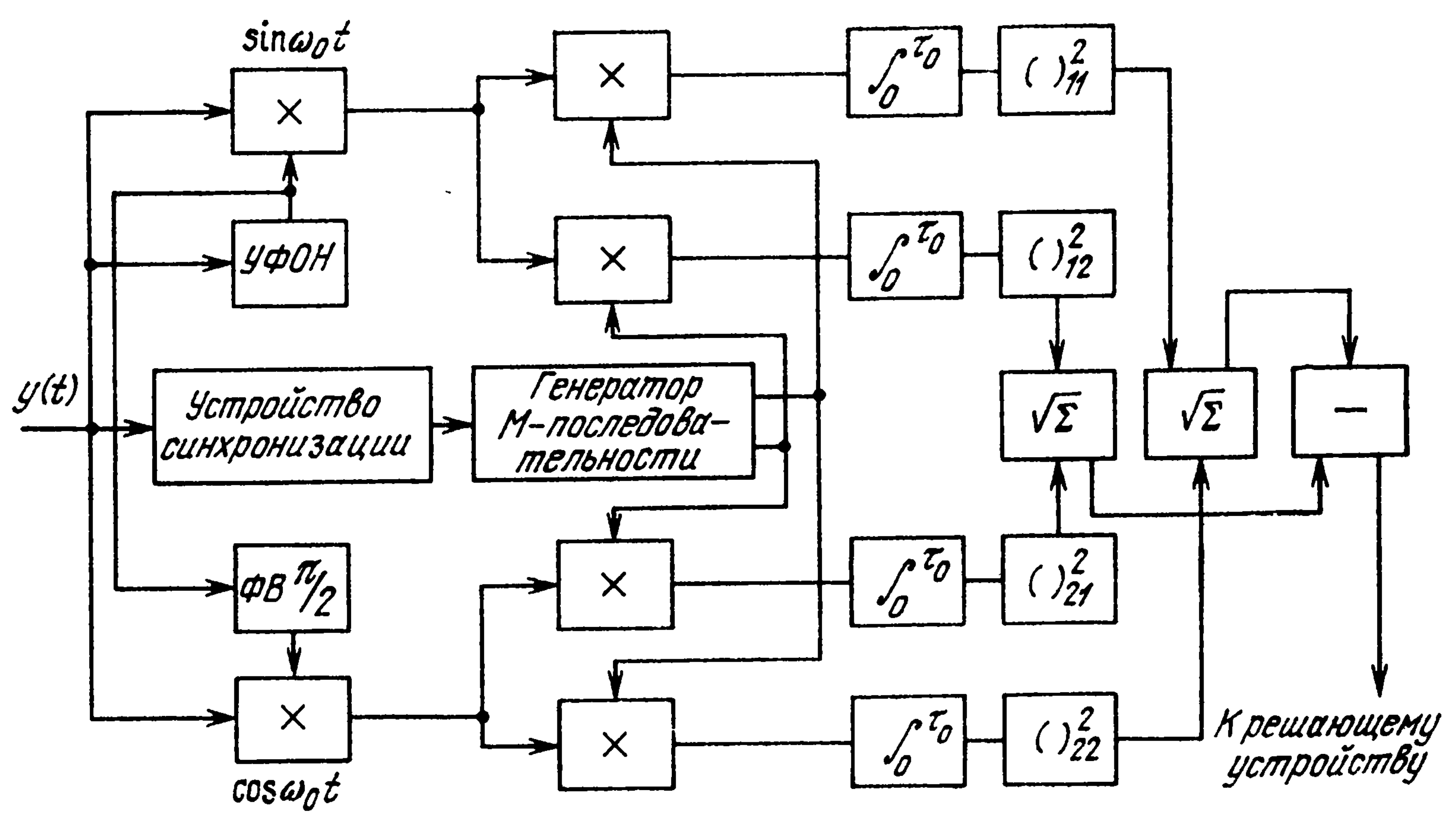


Рис. 6.13.

Схема приемника оказывается довольно сложной. Особенность ее работы по сравнению со схемой оптимального приема сложного сигнала, известного точно, состоит в том, что принятое колебание $y(t)$ разлагается на два ортогональных, соответствующих синусоидально-

му и косинусоидальному опорным напряжениям (рис. 6.14). При этом составляющие каждой проекции $y_s(t)$ и $y_c(t)$ оказываются когерентными с опорными напряжениями. Приемник определяет длину вектора каждой проекции, а затем, зная их, находит длину вектора принятого сигнала. Так как посылка сигнала может принять одно из двух возможных значений, то необходимо вычислять два возможных значения век-

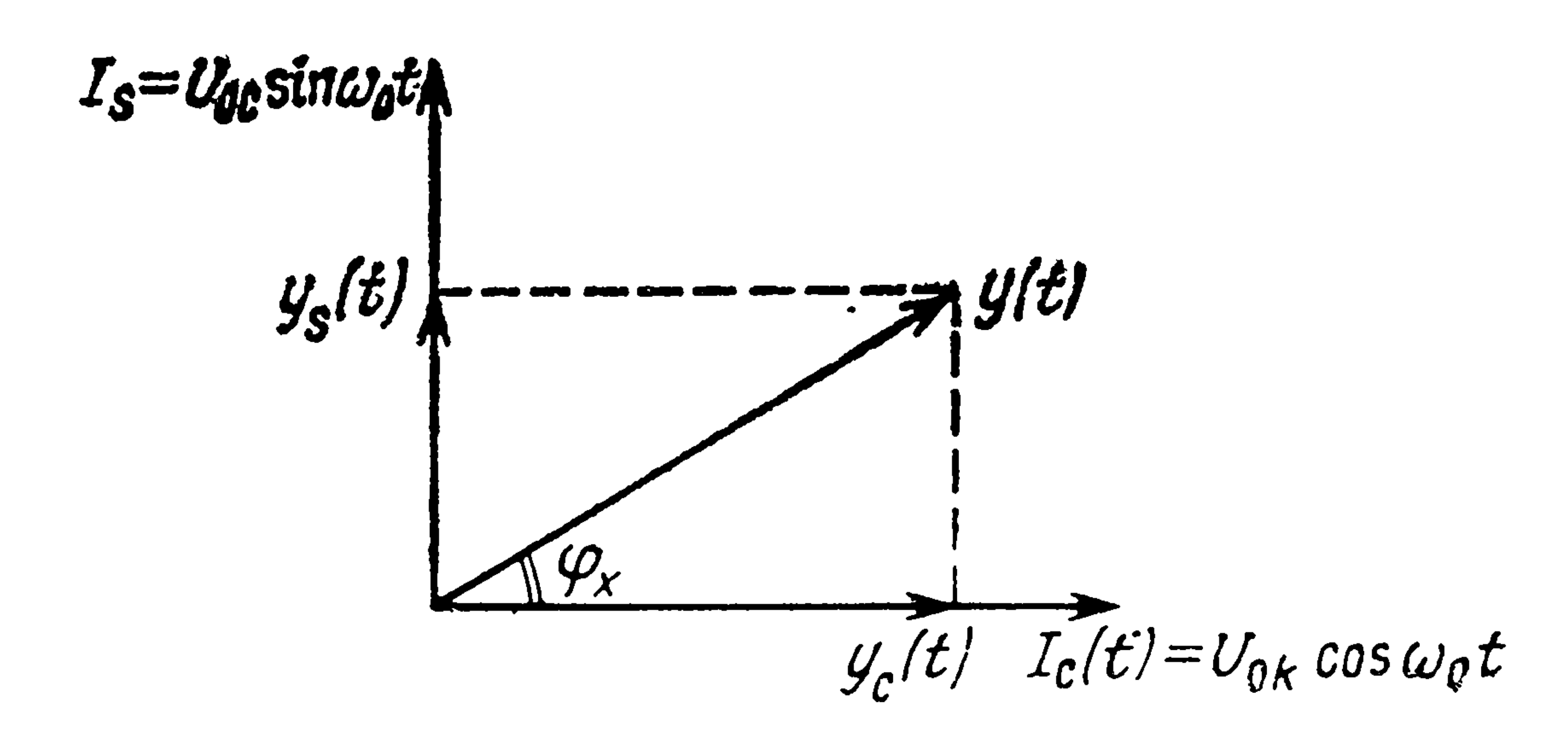


Рис. 6.14.

тора. Решение принимается после сравнения этих значений. Переданным считается тот информационный символ, которому соответствует принятый сигнал с большей величиной вектора.

При практической реализации приема сложных сигналов возникает ряд трудностей, которые приводят к ухудшению показателей работы системы по сравнению с оптимальным приемом. Эти трудности обусловлены неидеальностью характеристик элементов и узлов реальных приемных устройств, нестабильностью этих характеристик во времени из-за влияния различных дестабилизирующих факторов (случайных изменений температуры, питающих напряжений и т. п.), а также наличием технологического разброса параметров. К перечисленным причинам добавляются такие факторы, как неидеальность синхронизации, невозможность иметь точное значение несущей частоты в пункте приема и т. п.

Указанные обстоятельства приводят к тому, что даже сравнительно небольшие отклонения от оптимальных методов приема могут существенно ухудшить помехоустойчивость системы, особенно если база сигнала значительна. Все это делает проблему реализации при-

емных устройств сложных сигналов очень трудной и многообразной. Детальное рассмотрение некоторых вопросов указанной проблемы содержится в работе [9].

6.7. ПРИМЕРЫ ПРИМЕНЕНИЯ СЛОЖНЫХ СИГНАЛОВ В ЦИФРОВЫХ СИСТЕМАХ СВЯЗИ

Применение сложных сигналов позволяет в ряде случаев решать задачи передачи информации более успешно по сравнению с традиционными методами. К таким случаям можно отнести передачу информации по каналам с многолучевым эффектом распространения; ослабление мешающего действия различных систем, работающих в том же диапазоне частот; повышение эффективности использования выделенного диапазона частот; ослабление влияния узкополосных помех и т. п. Кроме этого, сложные сигналы открывают ряд возможностей построения систем с такими свойствами, которые не могут быть получены в системах с простыми сигналами: скрытность, крипто- и имитостойкость связи; возможность одновременно с передачей информации измерять координаты движущегося объекта, на котором расположена передающая часть системы, и т. д.

Ниже кратко рассматриваются некоторые примеры применения сложных сигналов.

6.7.1. Применение сложных сигналов в многолучевых каналах

Эффект многолучевого распространения в каналах со случайными параметрами приводит к появлению интерференции между соседними посылками сигнала и их замираниям. При простых сигналах влияние интерференции можно уменьшить, только увеличивая длительность посылок, что, естественно, приводит к уменьшению скорости передачи. Для ослабления существенных замираний уровня простые сигналы приходится принимать методами разнесенного приема.

Применение сложных сигналов в ряде случаев позволяет решить задачу ослабления влияния интерференции и замираний более успешно. Рассмотрим для этого рис. 6.15, на котором показаны огибающие одной посылки сложного сигнала, пришедшей по трем разным лучам. Полагаем, что замирания являются общими и за время, равное длительности посылки τ_0, амплитуда и начальная фаза элементов сигнала не меняются.

Оптимальная обработка посылки сложного сигнала корреляторами или согласованными фильтрами дает выходное напряжение, вид которого совпадает с корреляционной функцией этой посылки. Выходные напряжения, соответствующие одной и той же посылке и разным лучам, показаны на рис. 6.15,*б*. Длительность основного лепестка этого напряжения приблизительно равна величине

$$\tau_{\text{вых}} \approx 1/\Delta f_э = \tau_0/\Delta f_э \tau_0 = \tau_0/Б_с. \qquad (6.47)$$

Если база сигнала значительна ($Б_с \gg 1$), то $\tau_{вых} \ll \tau_0$ и, следовательно, в результате обработки посылки происходит ее «сжатие» по времени. Нетрудно видеть, что при условии

$$\Delta\tau_{з\,min} \geqslant \tau_{вых} \qquad (6.48)$$

выходные напряжения разделяются друг от друга несмотря на то, что на входе посылки от разных лучей перекрываются. Из (6.48) и (6.47) следует, что база сигнала, при которой обеспечивается разделение лучей с минимальной разностью запаздывания $\Delta\tau_{з\,min}$, должна удовлетворять условию

$$Б_с \geqslant \tau_0/\Delta\tau_{з\,min}. \qquad (6.49)$$

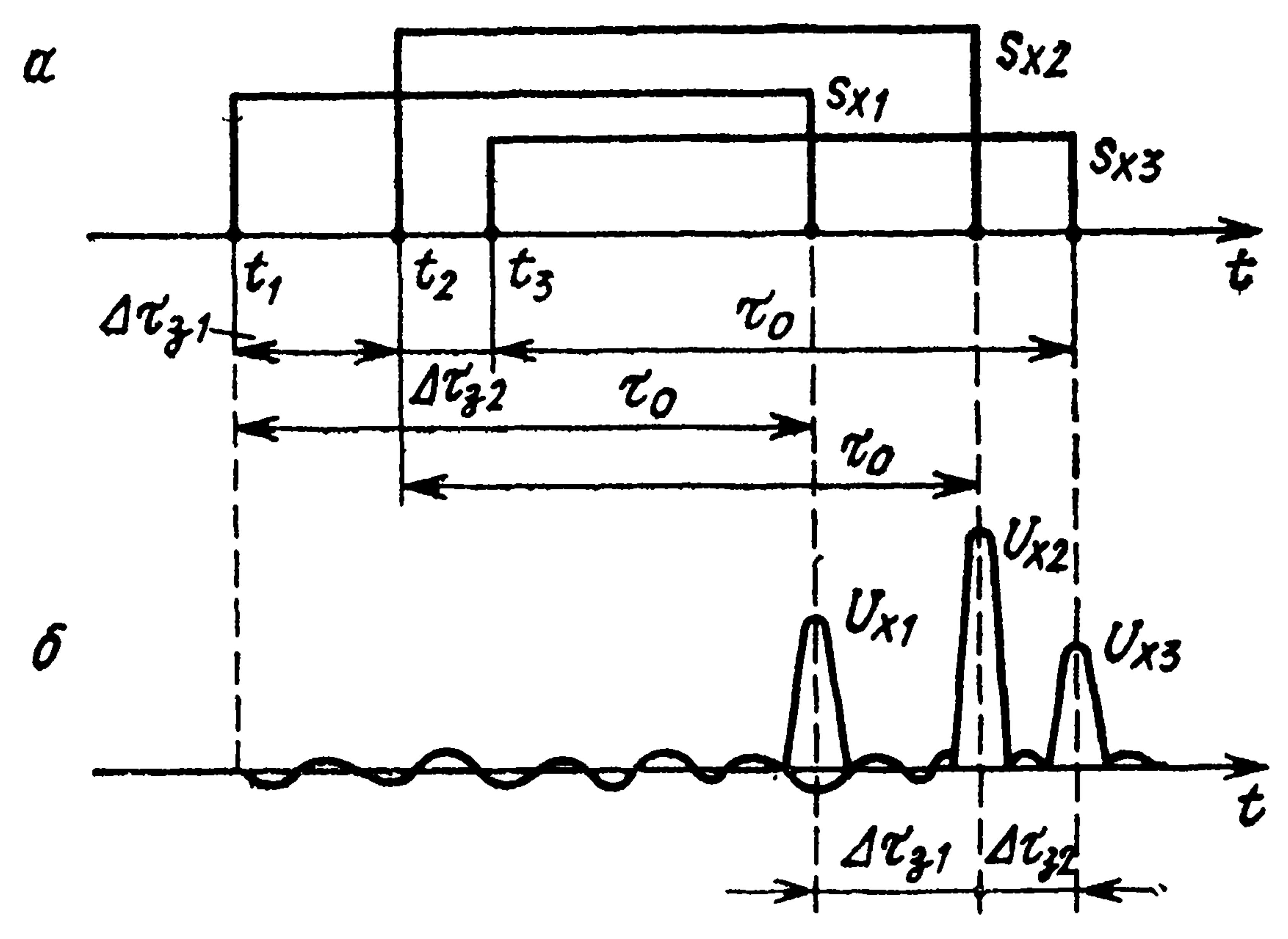


Рис. 6.15.

Так как огибающие посылки, пришедшей по разным лучам S_{x1}, S_{x2}, S_{x3}, случайны, то случайны и выходные напряжения U_{x1}, U_{x2}, U_{x3}.

Оптимальная обработка посылок представляет собой линейную операцию, поэтому закон распределения выходных напряжений остается таким же, как и у огибающих Если выходные напряжения обработать далее схемой, показанной на рис. 6.16,*а*, то на ее выходе образуется последовательность импульсов со случайными амплитудами (рис. 6.16,*б*), среди которых будет импульс с амплитудой $U_{xр}$. Поскольку этот импульс представляет собой результат сложения трех импульсов со случайными амплитудами U_{x1}, U_{x2} и U_{x3}, закон распределения величины $U_{xр}$ изменится по сравнению с распределениями случайных величин U_{x1}, U_{x2} и U_{x3}. В частности, если величины U_{x1}, U_{x2} и U_{x3} подчиняются распределению Релея, то распределение $U_{xр}$ определяется композицией этих распределений и, следовательно, замирания величин $U_{xр}$ менее глубоки, чем замирания случайных величин U_{x1}, U_{x2} и U_{x3}. Нетрудно видеть, что указанная процедура эквивалентна разнесенному приему с линейным сложением трех ветвей

Так как время запаздывания лучей и разность хода между ними случайны, то в общем случае необходимо применять линию с временем задержки

$$\tau_{л\,з} = \Delta\tau_{з\,max} - \Delta\tau_{з\,min} = \tau_р,$$

равным времени многолучевого растяжения сигнала (см. гл. 5), а отводы брать через интервалы, равные минимальному времени запаздывания $\Delta\tau_{з\,min}$. Значения $\Delta\tau_{з\,max}$ и $\Delta\tau_{з\,min}$ определяются при экспериментальных исследованиях свойств конкретного канала.

Таким образом, применение сложных сигналов позволяет разделить перекрывающиеся сигналы, приходящие по разным лучам. Это дает возможность эффективно использовать энергию сигналов от отдельных лучей и существенно ослабить влияние замираний, обусловленных эффектом многолучевого распространения.

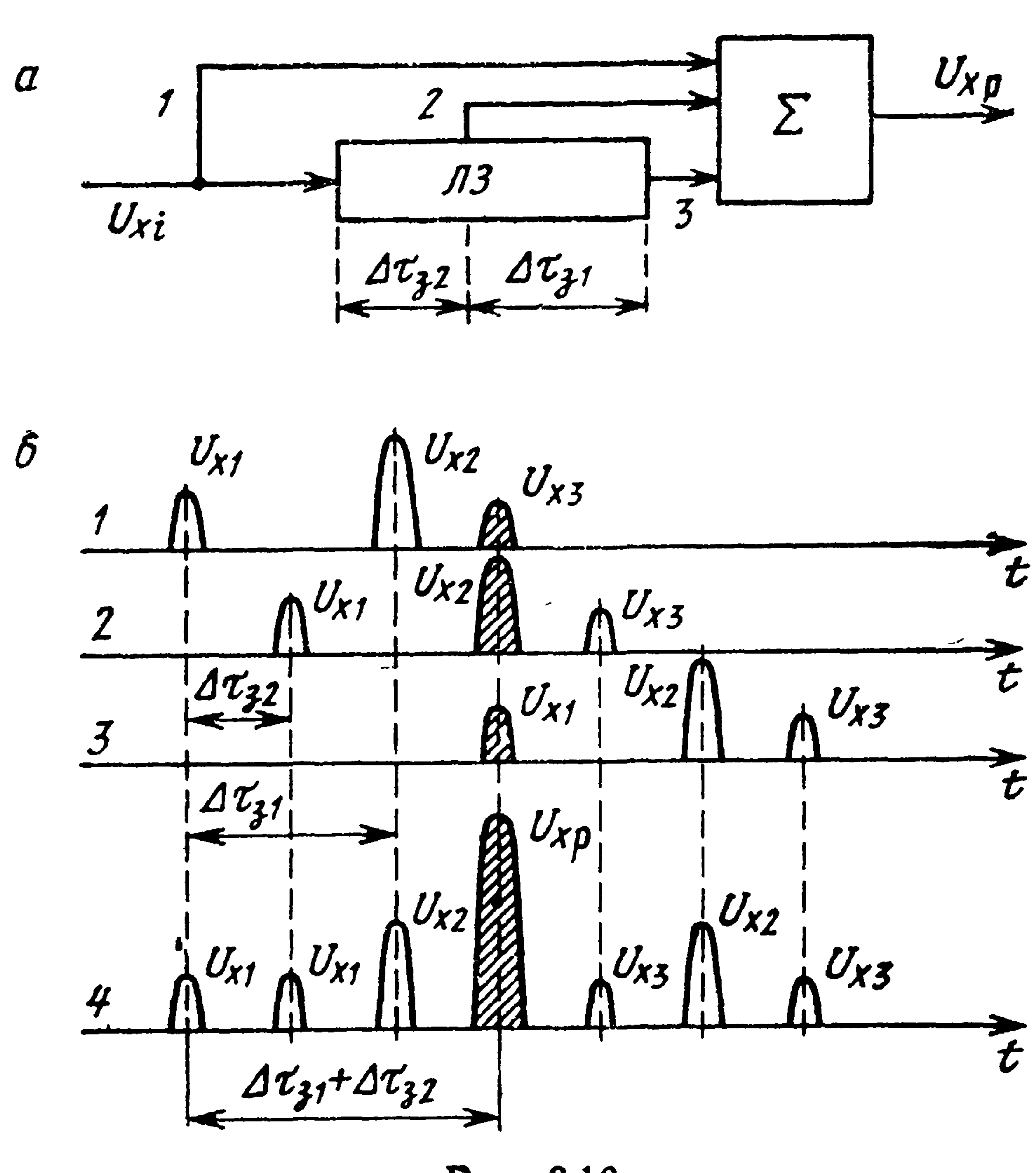


Рис. 6.16.

Изложенные выше идеи впервые были реализованы в 1956 г. в системе «Рейк» [18] *. Эта система предназначалась для передачи двоичной информации на расстояние около 4500 км методом ЧМн сигналов, манипулированных псевдослучайной бинарной последовательностью. Ширина спектра сигнала выбиралась равной 10 кГц, а база $Б_с=220$. В зависимости от времени суток использовались диапазоны 8, 12 и 17 МГц. При мощности передатчика 22 кВт система обеспечивала передачу информации с вероятностью искажения символа не более 10^{-6} и была эквивалентна системе примерно с 3—4-кратным пространственным разнесением.

---

* Название системы происходит от английского слова Rake — грабли. Действительно, напряжение на выходе сумматора (рис. 6.16) напоминает зубья грабель. Видимо, это и послужило основанием для такого термина.

Описание упрощенной структурной схемы приемника системы «Рейк» можно найти в работах [8, 18]. Система «Рейк» получила широкую известность и явилась прототипом ряда более сложных систем, некоторые из которых описаны в литературе [8].

6.7.2. Применение сложных сигналов для борьбы с помехами

Применение сложных сигналов иногда позволяет эффективно бороться с различными помехами, спектр которых сосредоточен в значительно более узком диапазоне частот по сравнению с диапазоном, занимаемым спектром сложного сигнала.

Как уже неоднократно указывалось, помехоустойчивость оптимального приема в гауссовских каналах зависит от величины (см. гл. 3).

$$h^2_0 = E_0/N_0 = (P_с/P_ш)_{вх}Б_с.$$

Для сложных сигналов с большим значением базы ($Б_с \gg 1$) требуемая величина h_0, при которой достигается заданная вероятность ошибки приема, может быть получена при малом входном отношении мощности сигнала к мощности шума. Это означает, что уровень сигнала на входе приемника оказывается значительно меньше уровня шума. Если ввести условную величину $N_{0с}$, характеризующую среднюю интенсивность мощности сигнала на единицу полосы, то для сигнала с большой базой можно записать условие

$$(P_с/P_ш)_{вх} = P_с/N_0\Delta f_э = (N_{0с}/N_0)_{вх} \ll 1.$$

Чем больше база сигнала, тем сильнее посылка «замаскирована» в шумах. При таких условиях обнаружить сам факт работы системы затруднительно. Так как мощность сигнала в такой системе распределена в широкой полосе частот, то ее интенсивность мала и такой сигнал практически не создает сколь-нибудь заметных помех системам, в которых используются значительно более узкополосные сигналы. Работа систем с узкополосными сигналами также не окажет существенного влияния на работу систем со сложными сигналами, если применять фильтрацию той части спектра сложного сигнала, где действует узкополосный сигнал. Потеря мощности сложного сигнала из-за такой фильтрации определяется соотношением

$$\Delta P_с/P_с = \Delta f_{у\,с}/\Delta f_э = 1/Б_с \ll 1,$$

где $\Delta f_{у\,с}$ — ширина спектра узкополосного сигнала. При большой базе эти потери незначительны и не приводят к заметному уменьшению величины напряжения после оптимальной обработки сложного сигнала.

Таким образом, применение сложных сигналов открывает возможность одновременной работы в одном и том же диапазоне частот узкополосных и широкополосных систем без заметного влияния их друг на друга (частотная совместимость систем).

6.7.3. *Применение сложных сигналов для улучшения эффективности использования отведенного диапазона частот*

Непрерывное увеличение числа различных систем связи делает особенно важной задачу рационального использования отведенных диапазонов частот. Наиболее распространенным до сих пор методом использования диапазона частот, выделенного для определенного класса систем, является метод частотной селекции сигналов. В соответствии с этим методом каждой системе выделяется определенный поддиапазон частот в пределах общего выделенного диапазона и разделение сигналов разных систем достигается их частотной селекцией.

Сложные сигналы можно разделять не только по частоте, но и по форме. Разделение по форме (кодовое или структурное разделение) основано на корреляционных свойствах ансамбля сложных сигналов, различающихся тонкой структурой (формой), обусловленной расположением последовательности элементов, образующих эти сигналы. Хорошие корреляционные свойства ансамбля сложных сигналов позволяют успешно использовать такие сигналы в системах, работающих одновременно в одном и том же диапазоне частот.

Выясним, при каких условиях такое применение сложных сигналов может улучшить эффективность использования отведенного диапазона частот по сравнению со случаем частотной селекции.

Пусть в отведенном диапазоне частот $\Delta f_{\text{общ}}$ необходимо организовать разветвленную сеть связи, позволяющую каждому из N абонентов в любое время связаться с любым другим абонентом.

При частотном уплотнении каждому абоненту выделяется полоса частот Δf_k, а для улучшения условий селекции этих полос вводятся защитные полосы $\Delta f_{\text{защ}}$. Тогда

$$\Delta f_k + \Delta f_{\text{защ}} = \Delta f_k (1 + \alpha_{\text{защ}}) = \Delta f_{\text{общ}}/N,$$

$$\Delta f_k = \Delta f_{\text{общ}}/N(1 + \alpha_{\text{защ}}),$$

где $\alpha_{\text{защ}}$ — коэффициент, учитывающий введение защитной полосы. При уплотнении по форме (кодовое или структурное уплотнение) все N абонентов используют весь отведенный диапазон.

Для сравнения эффективности использования полосы в первом и втором случаях воспользуемся критерием удельных затрат полосы. Тогда

$$\beta_{I\Delta f} = \Delta f_{\text{общ}} \Big/ \sum_{i=1}^{N} C_{Ii}, \quad \beta_{II\Delta f} = \Delta f_{\text{общ}} \Big/ \sum_{i=1}^{N} C_{IIi}, \tag{6.50}$$

где C_{Ii} и C_{IIi} — пропускная способность по i-му каналу в первом и втором случаях соответственно. Далее для простоты будем полагать, что эти значения не зависят от номера канала, и обозначим их C_{Ik} и C_{IIk}. Очевидно, что условием лучшей эффективности систем со сложными сигналами является

$$\beta_{II\Delta f}/\beta_{I\Delta f} = C_{Ik}/C_{IIk} < 1. \tag{6.51}$$

Для конкретизации этого условия воспользуемся формулой Шеннона. При этом будем считать, что мощность сигнала и спек-

тральная плотность шума в обоих случаях одинаковы. Тогда

$$C_{Ik}=\frac{\Delta f_{\text{общ}}}{N(1+\alpha_{\text{защ}})}\log_2\left[1+\frac{P_{\text{c}}N(1+\alpha_{\text{защ}})}{N_0\Delta f_{\text{общ}}}\right],$$

$$C_{IIk}=\Delta f_{\text{общ}}\log_2\left[1+\frac{P_{\text{c}}}{N_0\Delta f_{\text{общ}}}\right]. \qquad (6.52)$$

С учетом (6.52) условие (6.51) принимает вид

$$\frac{\log_2[1+AN(1+\alpha_{\text{защ}})]}{N(1+\alpha_{\text{защ}})\log_2(1+A)}<1, \qquad (6.53)$$

где $A=P_{\text{c}}/N_0\Delta f_{\text{общ}}=(P_{\text{c}}/P_{\text{ш}})_{\text{вх}}$.

Эта величина определяет отношение сигнал/шум на входе приемника системы со сложными сигналами. Для таких сигналов $A<1$. Так как отношение логарифмов не зависит от их основания, то, перейдя к натуральным логарифмам и использовав разложение $\ln(1+A)\approx A$ (при $A\leqslant 0{,}2$), получим

$$\frac{\ln[1+AN(1+\alpha_{\text{защ}})]}{AN(1+\alpha_{\text{защ}})}<1. \qquad (6.54)$$

Исследование этого условия показывает, что оно выполняется, начиная со значений

$$AN(1+\alpha_{\text{защ}})\geqslant 0{,}23. \qquad (6.55)$$

Величину $\alpha_{\text{защ}}$ обычно выбирают в пределах 0,1—0,3. Принимая $\alpha_{\text{защ}}=0{,}2$ и учитывая выражение для А, получаем из (6.55) следующее условие:

$$N(P_{\text{c}}/P_{\text{ш}})_{\text{вх}}\geqslant 0{,}2. \qquad (6.56)$$

Это условие определяет минимальное число каналов, начиная с которого система со сложными сигналами эффективнее использует полосу, чем системы с простыми сигналами. Чем меньше отношение сигнал/шум (чем больше база сигнала), тем при большем числе каналов начинается улучшение.

Примерами систем, эффективно использующих выделенный диапазон частот, могут служить различные дискретно-адресные системы с вызовом произвольного абонента, спутниковые системы связи со свободным доступом и т. п. Такие системы позволяют осуществлять связь между большим числом различных абонентов в любое необходимое время, т. е. эти системы являются несинхронными. Очень часто подобные системы называют асинхронно-адресными [13, 14].

Так как асинхронно-адресные системы, работающие в одном и том же диапазоне частот, могут создавать друг другу взаимные помехи, одной из основных проблем при их разработке является проблема выбора сигналов. Эти сигналы должны принадлежать к ансамблю с «хорошими» корреляционными свойствами (обеспечивать малые значения взаимокорреляционных функций по сравнению с основным пиком автокорреляционных функций). Для этого часто используют многочастотные составные сигналы, получаемые частотно-временным кодированием. Применение таких сигналов позволяет упростить приемно-передающую аппаратуру систем.

Обычно передачу цифровой информации в асинхронно-адресных системах ведут двоичными сигналами. Поэтому из выбранного

ансамбля каждой системе выделяется один (при передаче с пассивной паузой) или два (при передаче с активной паузой) сложных сигнала. Эти же сигналы одновременно выполняют роль «адреса» данной системы. Каждое приемное устройство может быть настроено на «адрес» любой системы, с которой необходимо установить связь.

Как уже указывалось в § 6.5, получить ансамбль большого числа сигналов, обладающих «хорошими» корреляционными свойствами, затруднительно. Поэтому некоторые сигналы могут давать заметные значения взаимокорреляционных функций. Если число одновременно работающих систем достаточно велико, то уровень помех на выходе коррелятора или оптимального фильтра приемника может быть значительным, так как он пропорционален сумме взаимокорреляционных функций сигналов работающих систем. Такие взаимные помехи являются основным фактором, ухудшающим качество передачи информации в адресных системах. Для ослабления их в подобных системах часто применяют определенную дисциплину работы. Если число активных систем возрастает настолько, что уровень взаимных помех становится недопустимым, то определенная часть систем должна прекращать работу.

С рядом практических применений асинхронно-адресных систем и примерами таких систем можно подробнее ознакомиться в работах [13, 14, 18, 19].

6.8. ОБЩИЕ СВЕДЕНИЯ О КОРРЕКТИРУЮЩИХ КОДАХ

В гл. 2 было подробно рассмотрено представление различных сообщений в цифровой форме с помощью первичных кодов. Одной из особенностей такого представления является его малая изобыточность или даже отсутствие избыточности. Действительно, если используются все кодовые комбинации первичного кода, то такой код оказывается полностью безызбыточным. Безызбыточный код обладает большой «чувствительностью» к помехам. Достаточно принять ошибочно только один символ, чтобы вместо переданной кодовой комбинации воспроизвести при приеме совсем другую комбинацию. Например, если при приеме комбинации 0100 произойдет ошибка в первой позиции, то принятой окажется комбинация 1100.

Таким образом, ошибка в любой из позиций безызбыточного кода приводит к замене одной кодовой комбинации другой. Это объясняется тем, что в безызбыточном коде отдельные комбинации могут отличаться друг от друга только в одной позиции (в одном разряде). Если для передачи сообщений использовать не все комбинации первичного кода, а только те, которые от-

личаются друг от друга не менее чем в двух позициях, то одиночная ошибка при приеме переведет используемую (разрешенную) комбинацию в неиспользуемую (запрещенную). Это позволяет обнаружить одиночную ошибку, которая может появиться при действии помех при приеме. В рассматриваемом случае одна разрешенная комбинация может перейти в другую разрешенную (необнаруживаемая ошибка) только при двойной ошибке. Вероятность такого события значительно меньше вероятности одиночной ошибки.

Итак, применяя для передачи цифровой информации код, у которого используются не все возможные комбинации, а только некоторые из них, можно повысить помехоустойчивость приема. Такие коды называют избыточными, или корректирующими. Корректирующие свойства избыточных кодов зависят от правила построения этих кодов, определяющего их структуру, и параметров кода (длительности символов, числа разрядов, избыточности и т. п.).

Простейшим примером корректирующего кода является код с проверкой на четность. Такой код строят следующим образом. К кодовым комбинациям безызбыточного первичного двоичного k-разрядного кода добавляется дополнительный разряд (позиция), называемый проверочным, или контрольным. Если число символов 1 в исходной кодовой комбинации четное, то в дополнительном разряде формируют контрольный символ 0, если число символов 1 нечетное, то в дополнительном разряде формируют контрольный символ 1. В результате общее число символов 1 в любой кодовой комбинации всегда должно быть четным.

Нетрудно видеть, что добавление дополнительного разряда увеличило общее число возможных комбинаций вдвое по сравнению с числом комбинаций исходного первичного кода, а условие четности разделило все комбинации на разрешенные и неразрешенные.

Код с проверкой на четность позволяет обнаруживать одиночную ошибку при приеме кодовой комбинации, так как такая ошибка нарушает условие четности, т. е. переводит разрешенную комбинацию в неразрешенную.

Увеличивая число дополнительных разрядов и формируя по определенным правилам проверочные символы 0 или 1, соответствующие этим разрядам, можно

усилить корректирующие свойства кода так, чтобы код позволял не только обнаруживать, но и исправлять ошибки. В связи с этим избыточные или корректирующие коды иногда разделяют на коды, обнаруживающие ошибки, и коды, исправляющие ошибки. Для обнаружения ошибки достаточно установить факт, что в данной кодовой комбинации произошла ошибка (одиночная, двойная и т. п.). Для исправления ошибки необходимо не только обнаружить ошибку, но и указать ту позицию в кодовой комбинации, где эта ошибка произошла. Естественно, что такая задача значительно сложнее, чем обнаружение ошибки, и требует применения более сложных кодов.

6.8.1. Основные виды корректирующих кодов

Исследованию корректирующих кодов посвящено огромное количество работ, начиная от сугубо математических и кончая отдельными частными работами прикладного характера. Поэтому в этой области, более чем в любой другой, существует множество различных терминов, часто выражающих одно и то же понятие. Известно большое число кодов, названных по фамилиям авторов, впервые предложивших эти коды. Все это затрудняет пользование литературой по вопросам кодирования, особенно для читателей, начинающих впервые знакомиться с этой обширной областью. Наиболее систематично вопросы кодирования изложены в работах [11, 20—24].

В настоящее время наибольшее внимание с точки зрения технических приложений уделяется двоичным равномерным корректирующим кодам, так как они обладают хорошими корректирующими свойствами и их реализация сравнительно проста. Двоичные равномерные коды делятся на блочные и непрерывные. Остановимся кратко на рассмотрении некоторых особенностей этих классов кодов.

Блочные коды. При использовании блочных кодов цифровая информация передается в виде отдельных кодовых комбинаций (блоков) равной длины. Кодирование и декодирование каждого блока осуществляется независимо друг от друга.

Почти все блочные коды относятся к разделимым кодам, т. е. таким, кодовые комбинации которых состоят

из двух различающихся частей: информационной и проверочной. Информационные и проверочные разряды во всех кодовых комбинациях разделимого кода всегда занимают одни и те же позиции. Разделимые коды обычно условно обозначают в виде (n, k), где n указывает общую значность кода (общее число позиции в блоке), k — число информационных позиций. Иными словами, величина n определяет общее число символов в блоке, а k — только число информационных символов*. Таким образом, число проверочных символов в разделимых кодах равно $r=n-k$.

Среди разделимых кодов различают систематические и несистематические коды.

Систематические, или линейные коды образуют наиболее обширную группу разделимых кодов. Особенностью систематических кодов является то, что проверочные символы образуются различными линейными комбинациями информационных символов. Теоретической основой получения таких комбинаций является аппарат линейной алгебры, который позволяет формировать проверочные символы по вполне определенным правилам (по определенной системе — отсюда и происхождение термина «систематические» коды). Использование аппарата линейной алгебры, в которой очень важное значение имеет понятие «группа», привело к тому, что в ряде работ рассматриваемые коды называют также «алгебраическими» или «групповыми» [20, 24].

Систематические коды обстоятельно изучены. Наиболее известны среди них циклические коды. Важными представителями циклических кодов являются коды Хэмминга, которые исторически появились раньше многих других кодов и сыграли большую роль в развитии теории корректирующих кодов, а также коды Боуза — Чоудхури. Для таких кодов найдены сравнительно простые методы их реализации. Исследования показали, что эти коды имеют особенно важное значение для приложений. Основное свойство циклических кодов состоит в том, что циклический сдвиг любой разрешенной кодо-

* Заметим, что здесь мы отступили от обозначений, использованных в предыдущих главах (m-значность кода, n — основание кода). Дело в том, что обозначение разделимого кода в виде (n, k) широко распространено в литературе и мы его сохранили в привычной записи.

вой комбинации также является разрешенной комбинацией. Циклические коды обладают хорошими корректирующими свойствами, а реализация кодирующих и декодирующих устройств для таких кодов оказывается проще, чем для других систематических кодов.

Непрерывные коды. Такие коды характеризуются тем, что в них проверочные символы перемежаются с информационными, и нет четкого деления получающейся на выходе кодирующего устройства последовательности символов на отдельные кодовые комбинации. Формирование проверочных символов в непрерывных кодах ведется по рекуррентным правилам, поэтому непрерывные коды часто называют цепными, или рекуррентными.

Исследованию непрерывных кодов уделяется большое внимание, поскольку кодирование и декодирование таких кодов несколько проще, чем систематических. Важно отметить, что непрерывные коды могут иметь хорошие корректирующие свойства, особенно для ошибок, возникающих «пачками» [11, 20, 24].

Проблема помехоустойчивого кодирования представляет собой обширную область теоретических и прикладных исследований. Ее основными задачами являются: 1) отыскание кодов, эффективно исправляющих ошибки требуемого вида; 2) нахождение методов кодирования и способов их реализации; 3) нахождение методов декодирования и приемлемых способов их реализации.

Наиболее разработаны эти задачи применительно к систематическим кодам. Такие коды успешно применяются в вычислительной технике и различных автоматизированных цифровых устройствах. Некоторые из предложенных кодов оказались удобными и для применения в цифровых системах передачи информации.

В следующих трех параграфах кратко излагаются некоторые особенности систематических корректирующих кодов, методы их кодирования и декодирования, а также рассматриваются возможности применения таких кодов для повышения помехоустойчивости приема.

6.8.2. Основные параметры корректирующих кодов

Основными параметрами, характеризующими корректирующие свойства кодов являются избыточность

кода, кодовое расстояние, число обнаруживаемых или исправляемых ошибок. Рассмотрим суть этих параметров.

Избыточность корректирующего кода. Избыточностью корректирующего кода называют величину

$$w=r/n=(n-k)/n=1-k/n. \qquad (6.57)$$

Из (6.57) следует

$$k/n=1-w. \qquad (6.58)$$

Эта величина показывает, какую часть общего числа символов кодовой комбинации составляют информационные символы. В теории кодирования величину k/n принято называть скоростью передачи [11]. Необходимо иметь в виду, что величина k/n характеризует относительную скорость передачи информации. Если производительность источника информации равна H символов в секунду, то скорость передачи после кодирования этой информации окажется равной

$$R=Hk/n, \qquad (6.59)$$

поскольку в закодированной последовательности из каждых n символов только k символов являются информационными.

Если число ошибок, которые нужно обнаружить или исправить, значительно, необходимо иметь код с большим числом проверочных символов. Чтобы при этом скорость передачи оставалась достаточно высокой, необходимо в каждом кодовом блоке одновременно увеличивать как общее число символов, так и число информационных символов. При этом длительность кодовых блоков будет существенно возрастать, что приведет к задержке информации при передаче и приеме. Чем сложнее кодирование, тем длительнее временная задержка информации.

Кодовое расстояние. При изучении корректирующих кодов очень полезно понятие о кодовом расстоянии или расстоянии Хэмминга. Расстоянием d между двумя кодовыми комбинациями называют число позиций, в которых эти комбинации имеют разные символы. Например, расстояние между комбинациями 0001101 и 1001010 равно четырем ($d=4$). Расстояние между различными комбинациями некоторого конкретного кода может быть различным. Так, в частности, в безызбыточном первичном натуральном коде это расстояние для различных

комбинаций может различаться от единицы до величины m, равной значности кода. Особую важность для характеристики корректирующих свойств кода имеет минимальное расстояние d_{min} между кодовыми комбинациями. Это растояние называют кодовым, или хэмминговым.

В безызбыточном коде все комбинации являются разрешенными и, следовательно, его кодовое расстояние равно единице. Поэтому достаточно исказиться одному символу, чтобы вместо переданной комбинации была принята другая разрешенная комбинация. Чтобы код обладал корректирующими свойствами, необходимо ввести в него некоторую избыточность, которая обеспечивала бы минимальное расстояние между любыми двумя разрешенными комбинациями не менее двух.

Число обнаруживаемых или исправляемых ошибок. Кодовое расстояние является основным параметром, характеризующим корректирующие способности данного кода. Если код используется только для обнаружения ошибок кратностью a, то необходимо и достаточно, чтобы минимальное расстояние было равно *

$$d_{min} \geqslant a+1. \tag{6.60}$$

В этом случае никакая комбинация из a ошибок не может перевести одну разрешенную кодовую комбинацию в другую разрешенную. Таким образом, условие обнаружения всех ошибок кратностью a можно записать в виде

$$a_{обн} \leqslant d_{min}-1. \tag{6.61}$$

Чтобы можно было исправить все ошибки кратностью a и менее, необходимо иметь минимальное расстояние, удовлетворяющее условию

$$d_{min} \geqslant 2a+1. \tag{6.62}$$

В этом случае любая кодовая комбинация с числом ошибок a отличается от каждой разрешенной комбинации не менее чем в $a+1$ позициях. Если условие (6.62) не выполнено, возможен случай, когда ошибка кратности a исказит переданную комбинацию так, что она станет ближе к одной из разрешенных комбинаций, чем к переданной, или даже перейдет в другую разрешен-

* Кратностью ошибки a называют число позиций кодовой комбинации, в которых под действием помехи одни символы оказались замененными другими. Например, 0 на 1 или 1 на 0.

ную комбинацию. Итак, условие исправления всех ошибок кратностью не более a можно записать в виде

$$a_{\text{испр}} \leqslant (d_{\min}-1)/2. \qquad (6.63)$$

Корректирующие коды можно одновременно использовать и для обнаружения и для исправления ошибок. В работе [11] указывается, что минимальное расстояние, при котором можно исправить все ошибки кратностью a или меньше и одновременно обнаружить все ошибки кратностью b или меньше, определяется условием

$$d_{\min} \geqslant a+b+1, \qquad (6.64)$$

где, в свою очередь, всегда должно иметь место условие $b>a$.

6.8.3. Оценка корректирующих возможностей линейных кодов

Вопрос о минимально необходимой избыточности, при которой код обладает нужными корректирующими свойствами, является одним из важнейших в теории кодирования. Этот вопрос до сих пор не получил полного решения. В настоящее время известен лишь ряд верхних и нижних оценок (границ), которые устанавливают связь между максимально возможным минимальным расстоянием корректирующего кода и его избыточностью [11]. Чтобы дать представление о количественной стороне этих оценок, ограничимся кратким изложением двух из них только для двоичных корректирующих кодов.

Нижняя граница Варшамова—Гильберта. Для больших значений n указанная граница определяется асимптотическим соотношением [11]

$$\frac{r}{n} = 1 - \frac{k}{n} \sim H\left(\frac{d_{\min}-2}{n-1}\right), \qquad (6.65)$$

где $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$.

Из выражения (6.65) следует

$$r \geqslant nH\left(\frac{d_{\min}-2}{n-1}\right), \qquad (6.66)$$

$$\frac{k}{n} \leqslant 1 - H\left(\frac{d_{\min}-2}{n-1}\right). \qquad (6.67)$$

Условия (6.66) и (6.67) позволяют оценить необходимое количество проверочных символов r и относительную

скорость передачи k/n при заданных (или выбранных) значениях n и d_{min}.

Верхняя граница Хэмминга. Основываясь на результатах, изложенных в [11], эту границу для двоичного корректирующего кода можно записать в виде

$$\frac{r}{n} \geqslant H\left(\frac{d_{min}-1}{2n}\right). \qquad (6.68)$$

Отсюда следуют выражения для оценки числа проверочных символов и относительной скорости передачи:

$$r \geqslant nH\left(\frac{d_{min}-1}{2n}\right), \qquad (6.69)$$

$$\frac{k}{n} \leqslant 1 - H\left(\frac{d_{min}-1}{2n}\right). \qquad (6.70)$$

Соотношение (6.70) известно в литературе как верхняя граница Хэмминга. Кроме рассмотренной оценки, известна также оценка, называемая верхней границей Плоткина [20]. Для двоичного корректирующего кода граница Плоткина определяется выражением

$$r \geqslant 2(d_{min}-1) - \log_2 d_{min}, \qquad (6.71)$$

которое справедливо, если $n \geqslant 2d_{min}-1$. Для значений $d_{min}/n \leqslant 0{,}3$ разница между границей Хэмминга и границей Плоткина сравнительно невелика.

Заметим, что для некоторых частных случаев Хэмминг указал более простые соотношения, позволяющие определить необходимое число проверочных символов [25]:

$$r \geqslant \log_2(n+1), \quad \text{если} \quad d_{min}=3, \qquad (6.72)$$

$$r \geqslant \log_2 2n, \quad \text{ели} \quad d_{min}=4. \qquad (6.73)$$

Систематические коды с $d_{min}=3$ и 4 в литературе обычно называют кодами Хэмминга.

6.9. ВОЗМОЖНОСТИ ПОВЫШЕНИЯ ПОМЕХОУСТОЙЧИВОСТИ ПРИЕМА ЦИФРОВОЙ ИНФОРМАЦИИ С ПОМОЩЬЮ КОРРЕКТИРУЮЩИХ КОДОВ

При передаче информации простым безызбыточным кодом достоверность приема в основном определяется типом канала и видом помех, которые в нем действуют.

В ряде случаев получающаяся достоверность недостаточна. Одним из путей ее повышения является применение корректирующего кода. Выбор того или иного кода и его параметров зависит от конкретных условий решаемой задачи: требуемой достоверности приема, допустимой относительной скорости передачи, вида ошибок в канале и т. п.

В теории помехоустойчивого кодирования принято различать два вида основных ошибок: статистически независимые, т. е. некоррелированные ошибки; статистически зависимые, т. е. коррелированные ошибки (ошибки типа пачек или пакетов).

Известно большое число математических моделей, позволяющих описать закономерности возникновения ошибок в различных каналах передачи цифровой информации [26, 27]. Рассмотрение вопросов применения корректирующих кодов для повышения достоверности передачи цифровой информации с учетом различных моделей ошибок представляет сложную задачу, особенно, если ошибки возникают «пачками». При этом для совокупности определенных условий требуется определенный код; например, если код предназначен для исправления только одиночных или двойных ошибок, то его применение в каналах с замираниями бесполезно. Исследование подобных вопросов представляет собой большую самостоятельную проблему, которая выходит за рамки данного пособия.

Основная цель дальнейшего рассмотрения состоит в том, чтобы проиллюстрировать на простом примере возможность повышения помехоустойчивости с помощью корректирующего кода и выяснить, при каких условиях применение такого кода целесообразно. При рассмотрении будем полагать, что ошибки в канале независимы, а сам канал симметричен. Такие условия характерны для гауссовских каналов, в которых применяются посылки сигналов с одинаковыми энергиями.

При независимых ошибках симметричный канал можно полностью охарактеризовать единственным параметром — вероятностью ошибочного приема символа. Все соотношения, характеризующие распределение ошибок, при этом оказываются простыми. Однако оценка корректирующих возможностей кода в таком простом случае имеет приближенный характер, и ее нужно рассматривать лишь как ориентировочную.

Как известно, прием цифровой информации возможен или посимвольно или в целом [27]. В первом случае каждая посылка сигнала, соответствующая определенному символу кодовой комбинации, анализируется в приемном устройстве отдельно (посимвольный или поэлементный прием), а затем принимается решение, к какой из возможных кодовых комбинаций следует отнести полученную последовательность символов. Во втором случае в приемном устройстве анализируется сразу вся принятая кодовая комбинация (прием в целом). Такой анализ предполагает, что число оптимальных (или подоптимальных) фильтров или корреляторов в приемном устройстве должно быть равно числу используемых (разрешенных) кодовых комбинаций.

При посимвольном приеме условия целесообразности применения корректирующего кода более жестки, чем при приеме в целом, поскольку помехоустойчивость приема в первом случае ниже, чем во втором *. Посимвольный прием допускает сравнительно простую реализацию и поэтому, несмотря на более низкую помехоустойчивость, применяется значительно шире, чем прием в целом. С учетом сказанного далее ограничимся рассмотрением только посимвольного приема кодовых комбинаций. При этом кроме принятых выше допущений о симметричности канала и независимости ошибок будем считать, что посимвольный прием осуществляется оптимально.

Пусть по гауссовскому каналу передается цифровая информация с помощью безызбыточного кода, содержащего k информационных разрядов и с длительностью символов, равной τ_0. Вероятность ошибочного приема символа в этом случае определяется выражением

$$p_1=0{,}5-\Phi(\gamma_c h_0), \qquad (6.74)$$

где $h^2_0=E_0/N_0=P_c\tau_0/N_0$.

Применение корректирующего кода означает увеличение избыточности в используемых комбинациях. Допустим, что теперь каждая кодовая комбинация состоит из n символов. Здесь, в свою очередь, возможны два случая: 1) увеличение n производится при сохранении прежней длительности кодовой комбинации ($\tau_к=k\tau_0=$

* Заметим, что для безызбыточного кода посимвольный прием и прием в целом обеспечивают одинаковую помехоустойчивость [27].

$=\text{const}$); 2) увеличение n производится при сохранении прежней длительности символа ($\tau_0=\text{const}$). Нетрудно видеть, что в первом случае

$$h^2_{\text{ок}}=\frac{E_{\text{1к}}}{N_0}=\frac{k}{n}h^2_0 \tag{6.75}$$

и

$$p_{1\text{к}}=0{,}5-\Phi(\sqrt{k/n}\,\gamma_c h_0), \tag{6.76}$$

где $p_{1\text{к}}$ — вероятность ошибочного приема символа в кодированной последовательности.

Во втором случае

$$h_{0\text{к}}=h_0 \text{ и } p_{1\text{к}}=p_1. \tag{6.77}$$

Чтобы оценить возможность повышения помехоустойчивости с помощью корректирующего кода, найдем вероятность неправильного приема отдельного символа с учетом корректирующих свойств кода. Поскольку такой код позволяет обнаружить или исправить часть ошибок, то, очевидно, не все ошибки в приеме отдельных символов имеют значение. Пусть кратность обнаруживаемой или исправляемой ошибки равна a. Тогда вероятность того, что в кодовой комбинации, состоящей из n символов, каждый из которых может быть принят с вероятностью ошибки $p_{1\text{к}}$, произойдет ошибка на a позициях кода, равна

$$p_{a\text{ош}}=p^a_{1\text{к}}. \tag{6.78}$$

Вероятность того, что на остальных ($n-a$) позициях ошибки не будет, равна

$$p_{(n-a)\text{прав}}=(1-p_{1\text{к}})^{n-a}. \tag{6.79}$$

Вероятность совместного события, состоящего в том, что в данной кодовой комбинации на a позициях произошла ошибка, а на $n-a$ позициях ошибки нет, равна

$$p=p^a_{1\text{к}}(1-p_{1\text{к}})^{n-a}. \tag{6.80}$$

В кодовой комбинации, состоящей из n символов, ошибка на a позициях может произойти в самых различных сочетаниях этих позиций. Число таких сочетаний определяется выражением

$$C^a_n=n!/a!(n-a)!. \tag{6.81}$$

Следовательно, вероятность того, что в кодовой комбинации из n символов появится ошибка кратностью a

(в любом сочетании позиции), определяется выражением

$$p_a = C^a_n p^a_{1к} (1 - p_{1к})^{n-a}. \quad (6.82)$$

Это выражение можно рассматривать как вероятность обнаруживаемой или исправляемой ошибки, так как по условию было принято, что все ошибки кратностью a и менее могут быть обнаружены или исправлены.

Если при передаче произойдут ошибки кратностью $a+1$ и выше, то такие ошибки уже не будут обнаружены или исправлены*. Нетрудно видеть, что полная вероятность необнаруживаемых или неисправляемых ошибок определяется выражением

$$p_{ош} = \sum_{j=a+1}^{n} C^j_n p^j_{1к} (1 - p_{1к})^{n-j}, \quad (6.83)$$

где $j = a+1, a+2, \ldots, n$.

При малой вероятности ошибки выражение (6.83) можно приближенно заменить первым членом суммы:

$$p_{ош} \approx C_n^{a+1} p_{1к}^{a+1} (1 - p_{1к})^{n-a-1}. \quad (6.84)$$

Это выражение можно рассматривать как вероятность ошибочного приема кодовой комбинации с учетом корректирующих свойств кода.

Для оценки эффективности корректирующего кода удобно пользоваться эквивалентной вероятностью ошибочного приема одного символа такого кода

$$p_{1э} = \frac{p_{ош}}{n} = \frac{1}{n} C_n^{a+1} p_{1к}^{a+1} (1 - p_{1к})^{n-a-1}. \quad (6.85)$$

Сравнивая эту величину с вероятностью ошибки символа для безызбыточного кода, определяемой выражением (6.74), можно установить, при каких условиях применение корректирующего кода позволяет повысить помехоустойчивость приема. Естественно, что такое сопоставление должно проводиться при некоторых одинаковых условиях (например, при одинаковой полосе занимаемых частот или одинаковой длительности кодовой комбинации и т. п.).

---

* Вообще говоря, корректирующий код способен исправлять некоторые ошибки и более высокой кратности, чем a. Однако вес таких ошибок среди других возможных ошибок невелик, и это обстоятельство можно не принимать во внимание.

В качестве примера рассмотрим эффективность кода Хэмминга с кодовым расстоянием $d_{min}=3$. Такой код позволяет обнаружить двойную ошибку или исправить одиночную.

Из соотношения (6.72) следует, что

$$n=2^r-1, \tag{6.86}$$

где r — число проверочных символов.

Возьмем $r=3$. Так как $k=n-r$, то с учетом выражения (6.86) нетрудно установить, что код имеет в данном случае вид $(n, k)=(7, 4)$.

Пусть передача безызбыточным 4-разрядным кодом ($k=4$) в гауссовском канале с оптимальным приемником характеризуется вероятностью ошибки на символ, равной $p_1=10^{-2}$. В соответствии с формулой (6.74) и таблицей П1 (см. приложение) находим, что вероятность $p_1=10^{-2}$ достигается при величине $\gamma_c h_0 \approx 2{,}33$.

Будем считать, что избыточность вводится за счет сокращения длительности символа первичного кода (расширения используемой полосы частот). Тогда в соответствии с соотношениями (6.75) и (6.76) имеем

$$h^2_{0к}={}^4/_7 h^2_0 \text{ и } p_{1к}=0{,}5-\Phi(\sqrt{4/7}\cdot 2{,}33)\approx 3{,}75\cdot 10^{-2}.$$

Пусть код Хэмминга используется для обнаружения ошибок. Тогда $a_{обн}\leqslant 2$ и эквивалентная вероятность необнаруживаемой ошибки определяется выражением (6.85), которое в данном случае принимает вид $p_{1э}=$ $=(^1/_7)\,C^3_7 p^3_{1к}(1-p_{1к})^4$. Подставив сюда найденное значение вероятности $p_{1к}$, получим $p_{1э}\approx 2{,}3\cdot 10^{-4}<p_1$. Этот результат показывает, что в данных условиях код Хэмминга, используемый для обнаружения двойной ошибки, оказывается эффективным. Если этот же код применять для исправления одиночных ошибок, то расчеты показывают, что при тех же условиях его эффективность ниже (в этом случае $p_{1э}=3{,}5\cdot 10^{-3}$).

Такая особенность характерна и для других корректирующих кодов. При заданной избыточности эффективность корректирующего кода при обнаружении ошибок всегда значительно выше эффективности этого же кода, используемого для исправления ошибок.

При протяженных кодах ($n\gg 1$) расчеты по оценке эффективности кодов становятся громоздкими. Для ориентировочных оценок можно использовать соотноше-

ние, полученное в работе [27]:

$$k/n > 1/d_{\min}. \quad (6.87)$$

Если это условие не выполнено, то применение корректирующего кода приведет к уменьшению помехоустойчивости.

6.10. МЕТОДЫ КОДИРОВАНИЯ И ДЕКОДИРОВАНИЯ СИСТЕМАТИЧЕСКИХ КОДОВ

Кодирование систематического кода сводится к образованию необходимого числа проверочных (избыточных) символов, которые обеспечивают обнаружение или исправление ошибок заданного вида. Любую кодовую комбинацию безызбыточного двоичного кода, имеющего k информационных разрядов, можно представить в виде

$$x_{i1},\ x_{i2},\ \ldots,\ x_{ij},\ \ldots,\ x_{ik}, \quad (6.88)$$

где индекс i принимает одно из двух значений ($i=1,\ 2$), указывая тем самым, какой вид имеет данный символ (например, $x_{1j}=1$, $x_{2j}=0$).

Введем в код избыточность, добавив к каждой кодовой комбинации (6.88) по r проверочных символов. Задача заключается в том, чтобы сформировать проверочные символы наиболее простым способом и обеспечить при этом обнаружение или исправление ожидаемого вида ошибок. Это достигается тем, что необходимая совокупность проверочных символов

$$z_{i1},\ z_{i2},\ \ldots,\ z_{1\text{к}},\ \ldots,\ z_{ir} \quad (6.89)$$

формируется так, чтобы каждый из них представлял собой линейную комбинацию информационных символов. Для этого нужно образовать систему r линейных уравнений, каждое из которых имеет вид

$$z_{ik} = \sum_{j=1}^{k} c_{ij} x_{ij} \pmod 2. \quad (6.90)$$

Здесь c_{ij} — весовые коэффициенты, принимающие значение 0 или 1 в зависимости от того, должен ли участвовать данный информационный символ x_{ij} в образовании проверочного символа z_{ik} с порядковым номером k.

Правило формирования кода обычно выбирают так, чтобы при декодировании имелась возможность выполнять ряд проверок на четность для некоторых, определенным образом выбранных, подмножеств информационных и проверочных символов каждой кодовой комбинации. Анализируя результаты выполненных проверок, можно обнаружить или исправить ошибку ожидаемого вида.

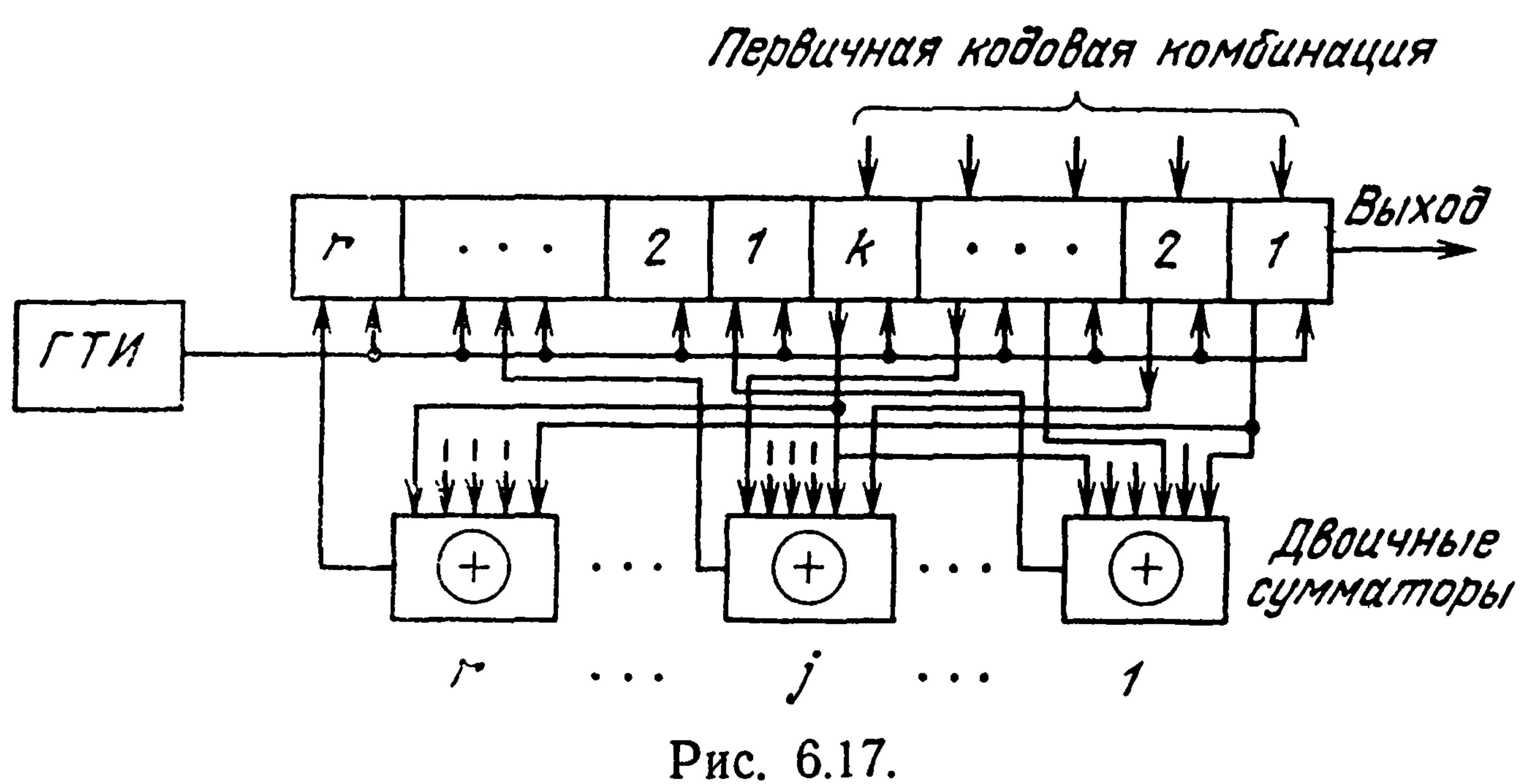


Рис. 6.17.

Правила построения систематических кодов довольно сложны, и мы на них останавливаться не будем. Подробное рассмотрение задач, связанных с построением корректирующих кодов, можно найти в работах [11, 22].

Как отмечалось выше, большим достоинством систематических кодов является сравнительная простота их технической реализации. Кодирующее устройство (n, k)-кода состоит из n разрядного регистра сдвига и r сумматоров по модулю 2 (рис. 6.17). Регистр содержит две части: информационную, состоящую из k ячеек, и проверочную, состоящую из r ячеек. Каждый сумматор служит для формирования проверочного символа, находящегося на определенной позиции. Подключение информационных ячеек регистра к соответствующим сумматорам выполняется в соответствии с правилом построения данного корректирующего кода.

Устройство работает следующим образом. Цифровая информация в виде кодовой комбинации первичного кода записывается параллельно (одновременно) в k информационных ячеек регистра. Одновременно с этим во всех r двоичных сумматорах происходит формирование проверочных символов, которые записываются в r

проверочных ячейках регистра. Следующим этапом является выведение полученной кодовой комбинации из регистра. Это достигается подачей тактовых импульсов (импульсов сдвига) от генератора ГТИ. После n тактовых импульсов кодовая комбинация окажется выведенной из регистра. При этом генератор ГТИ должен быть вновь отключен от регистра, а регистр подготавливается к записи и формированию следующей кодовой комбинации. Кодовая комбинация с выхода регистра поступает к модулятору передающей части системы связи.

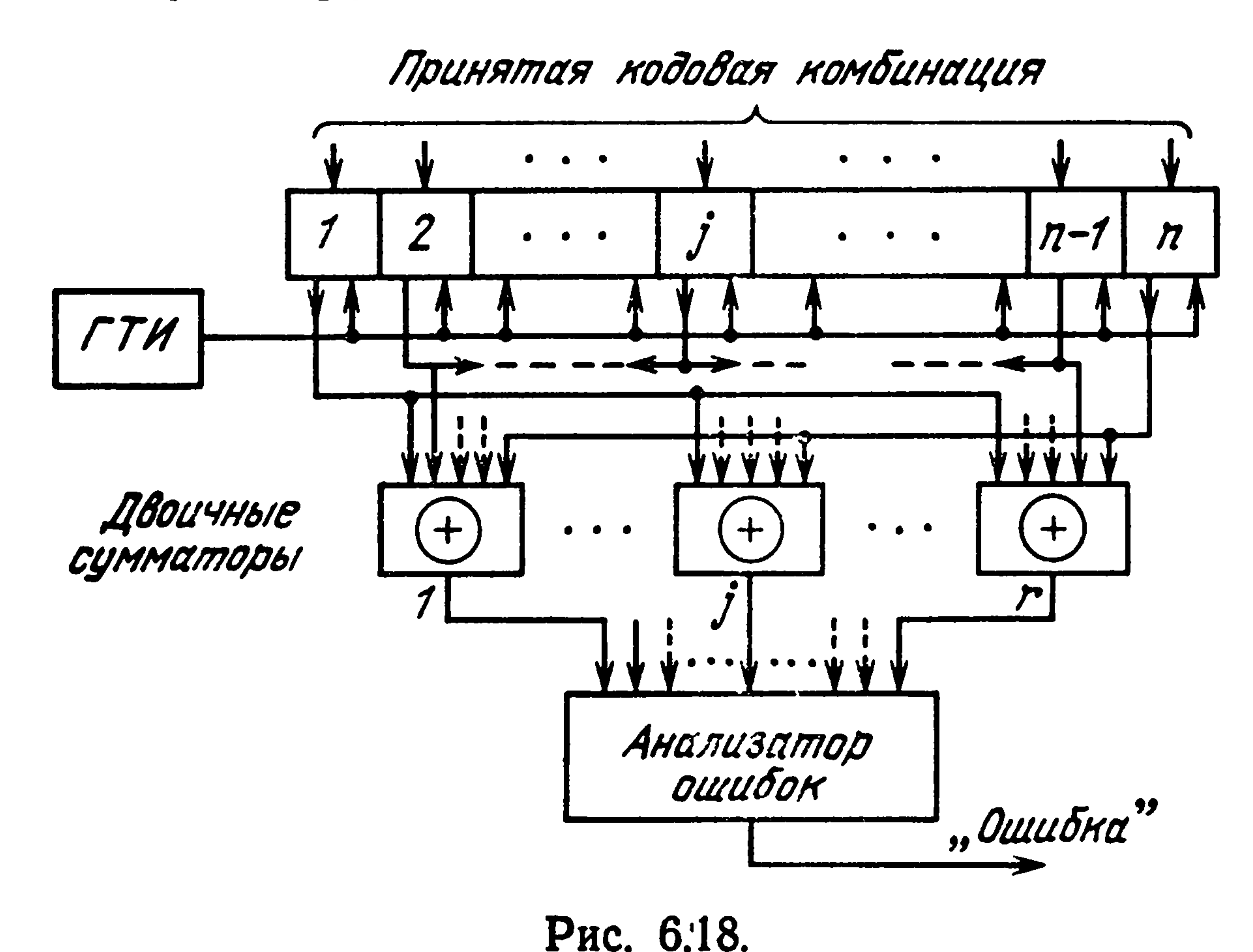


Рис. 6.18.

Формирование корректирующих кодов с помощью регистров сдвига получило очень широкое распространение. Этот метод прост в реализации, позволяет создавать схемы с высокой надежностью и обеспечивает возможность изменения параметров кода в процессе работы. Действительно, если изменить частоту следования тактовых импульсов, то изменится длительность каждого символа, а следовательно, изменится энергия кодовой комбинации, скорость передачи и т. д. Если при этом согласовано изменять соответствующие параметры в приемном устройстве, можно получить в некотором смысле адаптивную систему связи.

Рассмотрим принцип работы декодирующего устройства, упрощенная структурная схема которого показана на рис. 6.18. Устройство состоит из n триггерных ячеек, r сумматоров по модулю 2 и анализатора ошибок. По-

следовательность символов принятой кодовой комбинации (часть этих символов может быть принята неверно из-за действия помех) записывается в n ячеек регистра и с помощью сумматоров проводится проверка на четность для r подмножеств информационных и проверочных символов. Как уже говорилось выше, эти подмножества формируются при кодировании по определенным для данного кода правилам. Если в принятой кодовой комбинации ошибок нет, то на выходе всех сумматоров сформируются символы, соответствующие нулю. Если в принятой комбинации имеются ошибочные символы, то на выходе некоторых сумматоров сформируются символы, соответствующие единице, поскольку условия проверки на четность в этих сумматорах окажутся не выполненными. В этом случае анализатор ошибок выработает сигнал «ошибка».

Увеличивая избыточность кода и усложняя устройства проверки и анализа ошибок, можно обеспечить такие условия, при которых декодирующее устройство сможет не только обнаружить ошибки в приеме отдельных символов, но и исправить часть этих ошибок. Для этого схема декодирующего устройства должна быть существенно усложнена по сравнению с указанной на рис. 6.18. Схема должна содержать ряд дополнительных устройств, таких, как устройство памяти для запоминания исправляемой кодовой комбинации, логические устройства для исправления символов на тех позициях кодовой комбинации, где они оказались принятыми ошибочно, и т. п.

Многочисленные исследования показали, что для кодов, обладающих большими корректирующими возможностями (а следовательно, представляющих собой последовательности с большим числом символов n), число операций при оптимальном декодировании растет экспоненциально с увеличением числа избыточных символов. Это обстоятельство существенно усложняет задачу технической реализации декодирующих устройств. Поэтому в последние 10 лет было уделено большое внимание вопросам упрощения методов декодирования, чтобы зависимость числа необходимых операций при декодировании от длины кодовой последовательности была не экспоненциальной, а степенной или даже линейной [28—32]. В настоящее время разработаны два подхода к решению задачи упрощения процедуры декодирования: вероятностный и алгебраический.

Суть вероятностного, или последовательного декодирования состоит в том, что вся совокупность возможных кодовых комбинаций (включая разрешенные и запрещенные) разбивается на две группы: высоковероятные и маловероятные. Если принятая комбинация ближе к высоковероятной группе, она декодируется сразу же без проверки. Если же принятая комбинация ближе к маловероятным комбинациям, она декодируется с проверкой и исправлением ошибок. Иными словами, сильно искаженные комбинации исправляются, а слабо искаженные декодируются без исправления. Такое декодирование несколько уступает в помехоустойчивости оптимальному, но заметно проще в реализации [29,31].

Суть алгебраического декодирования состоит в том, что исправляется только определенная часть ошибок, т. е. корректирующие возможности кода используются только частично. Это означает, что используется неоптимальный алгоритм декодирования, допускающий более простую схемную реализацию. Особенно удобны с точки зрения простоты декодирования циклические коды, являющиеся одной из разновидностей систематических кодов. Для таких кодов разработаны эффективные методы алгебраического декодирования, которые получили название пороговых, или мажоритарных [28,30].

Проведенное краткое рассмотрение особенностей корректирующих кодов показывает, что их применение связано с существенным усложнением систем передачи информации, особенно из-за декодирующих устройств. Возникает вопрос, а насколько все это целесообразно и нужно? Действительно, до сих пор помехоустойчивое кодирование не нашло еще широкого применения в цифровых системах связи, за исключением некоторых космических систем и некоторых систем передачи данных. Однако прикладное значение проблемы кодирования с развитием различных автоматизированных систем, использующих ЦВМ, будет непрерывно возрастать. Стремление к повышению скорости передачи по существующим линиям связи и возрастающие требования к достоверности передаваемой информации неизбежно приведут к необходимости широкого применения корректирующих кодов.

6.11. ЦИФРОВЫЕ СИСТЕМЫ СВЯЗИ С ОБРАТНЫМ КАНАЛОМ

Рассмотренные в предыдущих параграфах методы повышения качества работы систем передачи цифровой информации основаны на введении некоторой заранее установленной избыточности в передаваемые сигналы. При многоосновных сигналах такая избыточность связана с увеличением необходимой мощности сигнала, при сложных сигналах — с увеличением их базы за счет расширения спектра сигнала, а при применении помехоустойчивых кодов — с усложнением структуры кодовых комбинаций, которое в конечном счете также эквивалентно расширению спектра используемых частот.

Реализация любого из указанных методов естественно связана с определенным усложнением систем, т. е. с увеличением аппаратурной избыточности. Применяя тот или иной метод, позволяющий обеспечить требуемые в данных условиях показатели работы системы связи, разработчик системы сразу платит за такие показатели определенную «цену» (необходимая избыточность сигнала, усложнение аппаратуры и т. п.). С этой точки зрения системы связи однонаправленного действия являются «жесткими», негибкими системами.

В последние годы большое внимание уделяется исследованиям и применению систем связи, в которых для повышения качества работы используется обратный канал. По этому каналу тем или иным способом передается информация, характеризующая условия приема в прямом канале. Используя эти сведения, можно существенно повысить достоверность передачи информации. Системы с обратным каналом по существу являются адаптивными системами, так как их характеристики и режим работы меняются в соответствии с состоянием канала связи и уровнем помех в нем.

Создание систем связи с обратным каналом в ряде случаев облегчается тем, что между двумя пунктами часто имеется двухсторонняя связь. Используя сравнительно небольшую часть пропускной способности канала в одном направлении, можно существенно повысить качество работы канала связи в противоположном направлении.

Системы с обратным каналом часто называют системами с обратной связью. Такие системы принято

разделять на два класса: системы с решающей обратной связью (системы с переспросом) и системы с информационной обратной связью (системы с подтверждением или квитированием).

Рассмотрим кратко особенности таких систем.

6.11.1. *Системы с решающей обратной связью*

Чтобы лучше уяснить особенности работы такой системы, рассмотрим простой случай, когда информация передается безызбыточным кодом и возможное повышение достоверности передачи достигается только благодаря обратному каналу.

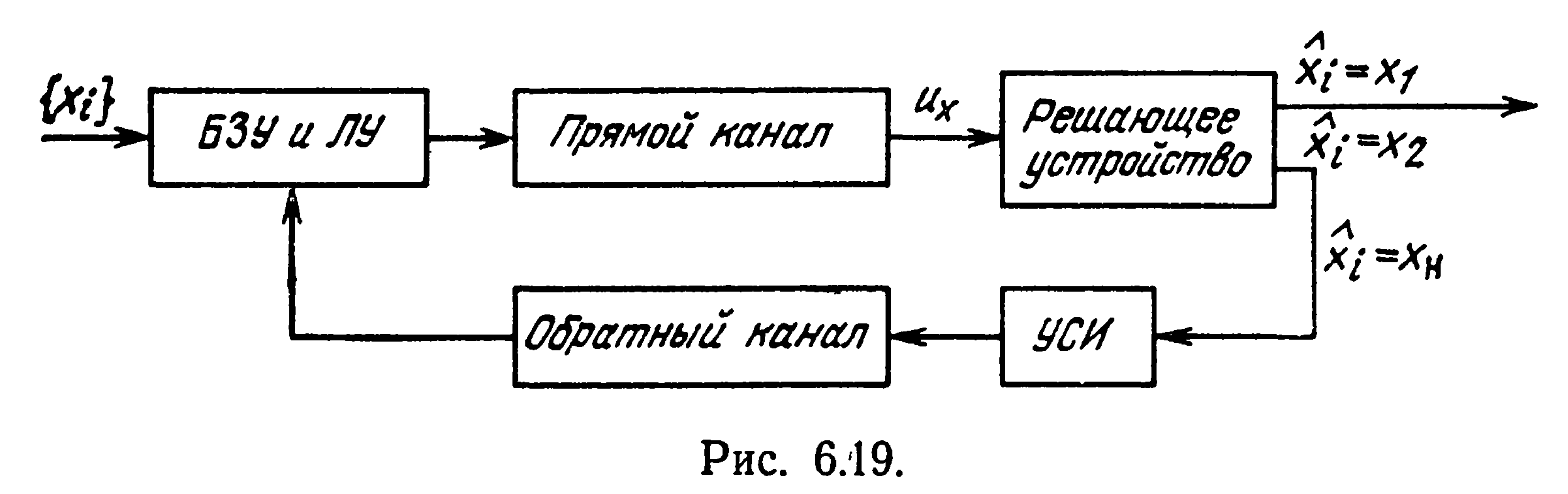


Рис. 6.19.

Упрощенная структурная схема системы в этом случае имеет вид, показанный на рис. 6.19 Модулятор и демодулятор системы на этом рисунке включены в состав прямого канала. Особенностью системы по сравнению с обычной двоичной цифровой системой является наличие буферного запоминающего устройства (БЗУ) и логического устройства (ЛУ) в передающей части системы, иное правило работы решающего устройства в приемной части системы, а также использование обратного канала.

Полагая, что прием ведется посимвольно, рассмотрим сначала работу решающего устройства*.

В отличие от решающих устройств двоичных цифровых систем связи оно имеет не один пороговый уровень

* Как известно, прием цифровой информации по кодовым комбинациям (прием «в целом») обеспечивает более высокую помехоустойчивость по сравнению с посимвольным приемом только при помехоустойчивом кодировании. Поэтому при безызбыточном кодировании прием обычно ведут посимвольно, так как он значительно проще и обеспечивает ту же помехоустойчивость, что и прием «в целом» [27].

$U_п$, а два (U_1 и U_2; $U_1 > U_2$), и решение принимается в соответствии с правилом

$$u_x > U_1 \quad \text{решение } \hat{x}_i = x_1,$$
$$u_x \leqslant U_2 \quad \text{решение } \hat{x}_i = x_2,$$
$$U_2 \leqslant u_x \leqslant U_1 \quad \text{решение не принято.} \qquad (6.91)$$

Здесь u_x — напряжение на выходе прямого канала (на выходе демодулятора); $\hat{x}_i$ — символ на выходе решающего устройства, соответствующий переданному символу x_i.

Правило (6. 91) разбивает все возможные решения на две области, разделенные зоной неопределенности, которая соответствует тем случаям, когда решение не принимается. Эту зону в литературе иногда называют зоной «стирания», так как при попадании напряжения u_x в эту зону переданный символ x_i не учитывается («стирается»). Прием с введением зоны стирания можно вести и в системах однонаправленного действия, однако он наиболее эффективен в системах с обратным каналом.

Таким образом, прием каждого символа связан с одним из трех возможных событий: 1) символ принят правильно; 2) символ принят ошибочно (вместо x_1 принят x_2 или наоборот); 3) решение не принято.

Если решение не принято, то в устройстве выработки служебной информации (УСИ) формируется символ переспроса, который передается по обратному каналу в передающую часть системы. Под действием этого переспроса логическое устройство обеспечивает повторную передачу информационного символа, который хранился в БЗУ. Если повторная передача того же символа опять не приведет к принятию решения на приемной стороне, то переспрос повторяется, а информационный символ передается третий раз и т. д.

Если решение на приемной стороне принято, сигнал переспроса не передается и система работает как обычная система однонаправленного действия до тех пор, пока не появится сигнал переспроса.

Исследования показывают [33], что в системах с переспросом весьма эффективно применение корректирующих кодов, обнаруживающих ошибки. Принятые кодовые комбинации разбиваются решающим устройством на две группы: разрешенные и неразрешенные комбинации. Если комбинация отнесена к разрешенной, то считается, что такая же комбинация была и передана. Естественно,

что имеется некоторая вероятность того, что это решение ошибочно, так как под действием помех одна разрешенная комбинация может быть переведена в другую. Однако эту вероятность необнаруживаемой ошибки можно сделать очень малой, выбирая соответствующий код. Если принятая комбинация отнесена к группе неразрешенных, то на приемной стороне формируется сигнал переспроса, по которому передающая часть повторяет передачу соответствующей кодовой группы.

Нетрудно видеть, что применение обнаруживающих кодов в системах с обратным каналом позволяет не только обнаруживать, но и исправлять ошибки, повторяя передачи тех комбинаций, которые были искажены помехами.

Исследования некоторых возможных вариантов систем с переспросом свидетельствуют о перспективности их применения для передачи цифровой информации [27, 33.] Экспериментальные исследования одной из подобных систем, получившей название «SECO», показали, что по стандартному телефонному каналу значительной протяженности (600—1200 км) можно передавать информацию со скоростью 6000—9000 бит/с при вероятности ошибки на символ, меньшей 10^{-9} [4, 34]. Эти показатели существенно выше, чем у односторонних линий связи. Однако необходимо отметить, что практическое применение систем с обратным каналом наталкивается на значительные технические трудности, связанные с большой сложностью аппаратуры [4, 23, 34].

6.11.2. Системы с информационной обратной связью

Особенностью таких систем является то, что каждое принятое цифровое сообщение может поступить к получателю только после определенной «проверки», выполняемой на передающей стороне связи. Для этого каждое принятое сообщение передается по обратному каналу в пункт передачи, где оно сравнивается с исходным сообщением. Если сообщения совпадают или различаются в допустимых пределах, то на передающей стороне принимается решение о том, что сообщение принято правильно и посылается сигнал подтверждения, в соответствии с которым принятое ранее сообщение, хранящееся в запоминающем устройстве, поступает к получа-

телю. Если же различие между сообщениями превышает допустимое, передатчик посылает сигнал о том, что принятое сообщение недостоверно, и повторяет передачу. Очевидно, что для этого на передающей стороне системы необходимо иметь запоминающее устройство, в котором исходное сообщение должно храниться до тех пор, пока оно не будет правильно принято в приемнике.

Сообщения, передаваемые в рассматриваемых системах по обратному каналу, часто называют «квитанциями» или «подтверждениями», поэтому и системы называют иногда системами с квитированием. Известно несколько разновидностей подобных систем [27, 35]. В частности, если для передачи применять избыточные коды, то по прямому каналу можно передавать только информационные символы, а по обратному — только проверочные, сформированные по принятым информационным символам. Сравнивая на передающей стороне проверочные символы, полученные по обратному каналу, с проверочными символами, хранящимися в запоминающем устройстве передатчика, можно установить степень их соответствия и принять необходимое решение.

Рассмотрение принципа действия систем с информационной обратной связью показывает, что такие системы целесообразно применять в тех случаях, когда скорость передачи не является фактором первостепенной важности и в первую очередь необходимо обеспечить высокую достоверность передаваемых сообщений (например, передача команд).

Ряд дополнительных сведений о таких системах можно найти в работах [23, 27, 35].

В заключение необходимо отметить, что теория систем с обратным каналом очень сложна и еще далека от завершения, а многие исследования выполнены с существенными идеализациями и допущениями. Разработка достаточно строгой теории таких систем представляла бы очень большой интерес не только для более успешного решения технических задач передачи информации, но и для решения ряда биологических задач, связанных с изучением механизмов наследственности, влияния среды и т. д.

6.12. КРАТКИЕ ИТОГИ ГЛ. 6

1. В этой главе были рассмотрены основные пути повышения качества работы систем передачи цифровой

информации. Рассмотрение показало, что выбор того или иного пути определяется задачами, которые должна решать система, условиями ее работы и существующими ограничениями технического и принципиального характера. Эти обстоятельства приводят к тому, что указанная проблема становится очень сложной и ее решение не может быть однозначным.

2. Все известные методы повышения качества работы систем связи основаны на использовании некоторой искусственно вводимой избыточности (аппаратурной, сигнальной, информационной или их сочетаний). Разнообразие видов избыточности открывает возможности выбора таких методов, которые наиболее целесообразны в данных конкретных условиях.

3. Эффективное использование отведенной полосы частот в каналах с малым уровнем помех и достаточно хорошим постоянством характеристик достигается применением многопозиционных простых сигналов с фазовой манипуляцией. Применение таких сигналов позволяет существенно повысить скорость передачи по каналам с ограниченной полосой пропускания. Однако стремление уменьшить затраты полосы наталкивается на значительное трудности, связанные с необходимостью существенного увеличения отношения сигнал/шум в канале и повышением требований к стабильности его характеристик. Эти обстоятельства обычно не позволяют применять на практике многопозиционные ФМн сигналы с кратностью больше трех.

4. Успешное решение ряда таких важных задач, как высокая помехоустойчивость к различным помехам, передача цифровой информации с высокой достоверностью по каналам со случайными параметрами, хорошая электромагнитная совместимость различных систем, использующих один и тот же диапазон частот, может быть достигнута применением сложных сигналов. Такие сигналы позволяют эффективно использовать энергию передатчика, но вместе с тем требуют значительного расширения используемого диапазона частот по сравнению с простыми сигналами.

5. Положительные свойства сложных сигналов проявляются наиболее полно только при условии, что способы их приема близки к оптимальным. Реализация таких способов приема встречает значительные трудности, которые резко возрастают при увеличении базы сигнала.

6. Эффективным методом ослабления действия различных помех является применение корректирующих кодов. Такие коды целесообразно применять в тех случаях, когда требуется обеспечить высокую достоверность передачи без существенного снижения ее скорости. Особенно важное значение это имеет для линий передачи данных, автоматизированных комплексов управления и т. п.

7. Основные трудности применения корректирующих кодов связаны с реализацией декодирующих устройств, так как их сложность резко возрастает с увеличением корректирующих возможностей кода. Однако разработанные к настоящему времени упрощенные методы декодирования и достижения современной микроэлектроники делают эту проблему технически разрешимой.

8. Большие возможности существенного повышения достоверности передачи информации открывает применение систем связи с обратным каналом. Применение в таких системах сравнительно несложных корректирующих кодов с обнаружением ошибок позволяет обеспечить такую достоверность передачи информации, которая не может быть достигнута в системах однонаправленного действия. Значительным достоинством систем с обратным каналом является их адаптация к изменениям условий работы.

9. Рассмотренные в этой главе методы повышения качества работы систем связи не универсальны, и каждый из них наряду с определенными достоинствами обладает и рядом недостатков. К числу таких недостатков можно отнести сложность реализации и достаточно высокую «чувствительность» к отклонениям условий приема от оптимальных. С этой точки зрения передача цифровой информации простыми двоичными сигналами наиболее выгодна — она проста в реализации и сравнительно малочувствительна к отличиям способов приема от оптимальных. Поэтому применять те или иные методы повышения качества передачи информации необходимо тогда, когда их преимущества очевидны и усложнение системы оправдано.

КОНТРОЛЬНЫЕ ВОПРОСЫ И ЗАДАЧИ

6.1. Указать основные пути повышения качества работы систем передачи цифровой информации и области их применения.

6.2. Дать характеристику основных особенностей многопозиционных сигналов, эффективно использующих отведенную полосу частот.

6.3. Дать физическое объяснение того, что многопозиционные сигналы с ФМн требуют меньшего отношения сигнал/шум, чем многопозиционные сигналы с АМн.

6.4. Показать, что выражение (6.12) эквивалентно выражению

$$(P_с/P_ш)_{вх} \geqslant 0{,}1kh^2_{отр} \exp(0{,}7k),$$

где k — кратность манипуляции.

6.5. Сравнить необходимое отношение мощности сигнала к мощности шума в многопозиционной системе с ФМн с соответствующим отношением в идеальной по Шеннону системе. Принять $M=8$, вероятность ошибки символа в системе с ФМн с оптимальным приемом $p_{ош}=10^{-5}$.

6.6. Указать причины, препятствующие применению многопозиционных ФМн сигналов с кратностью более трех.

6.7. Назвать способы приема сигналов с многократной ФМн и дать их характеристику.

6.8. Указать основные особенности системы «Кинеплекс». Почему в этой системе полоса частот разбивается на ряд поддиапазонов?

6.9. Каким требованиям должны удовлетворять сложные сигналы и в чем состоит суть этих требований?

6.10. Какими параметрами можно охарактеризовать сложный сигнал и его свойства?

6.11. В чем проявится различие между сложными сигналами, имеющими различные длительности и ширину спектра, но одинаковую базу? В чем проявится сходство?

6.12. Изложить основные особенности методов получения сложных сигналов.

6.13. Построить автокорреляционные функции последовательностей

010110100101011
011001010110101

Найти взаимокорреляционную функцию этих последовательностей. Чему равен коэффициент корреляции между последовательностями при отсутствии сдвига между ними?

6.14. Показать, что линейный регистр сдвига из четырех ячеек с начальными состояниями $\gamma_{01}=1$, $\gamma_{02}=0$, $\gamma_{03}=0$, $\gamma_{04}=1$ и обратной связью, формируемой по правилу $F_л=\gamma_3 \oplus \gamma_4$, генерирует M-последовательность вида

000100110101111

6.15. Изложить особенности оптимальных методов приема сложных сигналов и объяснить работу оптимальных схем, показанных на рис. 6.12 и 6.13.

6.16. Определить, какому приблизительно числу ветвей разнесенного приема соответствует система «Рейк», если при базе 220 она обеспечивала передачу двоичной информации с вероятностью ошибки $p_{ош} \leqslant 10^{-6}$, а при уменьшении базы до 50 вероятность ошибки возросла до 10^{-2}. Прием в обоих случаях считать оптимальным, канал — релеевским.

6.17. Оценить корректирующие способности систематического кода (15,5). Сколько и какой кратности ошибок можно исправить или обнаружить с помощью такого кода?

6.18. Показать, что для кодов Хэмминга справедливы следующие соотношения:

$$\frac{k}{n} \leqslant 1 - \frac{1}{n} \log_2 (n-1), \quad n \leqslant 2^r - 1, \quad d_{\min} = 3;$$

$$\frac{k}{n} \leqslant 1 - \frac{1}{n} \log_2 2n, \quad n \leqslant 2^{r-1}, \quad d_{\min} = 4.$$

6.19. Показать, что код Хэмминга при $d_{\min}=3$ может иметь ряд вариантов: (7, 4); (15, 11); (31, 26); (63, 57); (127, 120) и т. д. Обсудить, при каких условиях целесообразно применять те или иные варианты кода, имея в виду, что корректирующие возможности этих вариантов одинаковы.

6.20. Показать, что при сохранении энергии кодовой комбинации неизменной и увеличении избыточности кода вероятность ошибки имеет минимум при некотором значении $r=r_{\text{опт}}$.

6.21. В чем различие между системами с решающей и информационной связью? Укажите возможные области их применения.

6.22. Привести примеры, когда абоненты, ведя разговор по телефону, используют канал связи как канал с переспросом, а когда как канал с информационной обратной связью.

СПИСОК ЛИТЕРАТУРЫ

1. **Стейн С., Джонс Дж.** Принципы современной теории связи и их применение к передаче дискретных сообщений. М., «Связь», 1971.

2. **Петрович Н. Т.** Передача дискретной информации в каналах с фазовой манипуляцией. М., «Сов. радио», 1965.

3. **Заездный А. М., Окунев Ю. Б., Рахович Л. М.** Фазоразностная модуляция и ее применение для передачи дискретной информации. М., «Связь», 1967.

4. **Базилевич Е. В. и др.** Передача данных. М., «Связь», 1969.

5. **Бухвинер В. Е.** Дискретные схемы в фазовых системах радиосвязи. М., «Связь», 1969.

6. **Фазовая** и относительная фазовая телеграфия. Сб. статей. М., «Связь», 1967.

7. **Передача** цифровой информации. Пер. с англ. Под ред. С. И. Самойленко М., ИЛ, 1963.

8. **Петрович Н. Т., Размахнин М. К.** Системы связи с шумоподобными сигналами. М., «Сов. радио», 1969.

9. **Шумоподобные** сигналы в системах передачи информации. Под ред. В. Б. Пестрякова. М., «Сов. радио», 1973.

10. **Теория** и применение псевдослучайных сигналов. М., «Наука» 1969. Авт.: А. И. Алексеев, А. Г. Шереметьев, Г. И. Тузов, Б. И. Глазов.

11. **Питерсон У.** Коды, исправляющие ошибки. Пер. с англ. М., «Мир», 1964.

12. **Цифровые** методы в космической связи. Под ред. С. Голомба. Пер. с англ. М., «Связь», 1969.

13. **Венедиктов М. Д., Марков В. В., Эйдус Г. С.** Асинхронные адресные системы связи. М., «Связь», 1968.

14. **Соммер Р.** Высокоэффективная система связи с многократным доступом через ретранслятор с обработкой сигналов. — «Зарубежная радиоэлектроника», 1968, № 7.

15. **Бобнев М. П.** Генерирование случайных сигналов. М., «Энергия», 1971.

16. **Корн Г. А.** Моделирование случайных процессов на аналоговых и аналого-цифровых машинах. М., «Мир», 1968.

17. **Окунев Ю. Б., Яковлев Л. А.** Широкополосные системы связи с составными сигналами. М., «Связь», 1968.

18. Широкополосные системы связи. (Обзор работ). — «Зарубежная радиоэлектроника», 1965, № 8.

19. Широкополосные многоканальные системы связи. (Обзор работ). — «Зарубежная радиоэлектроника», 1967, № 10.

20. **Харкевич А. А.** Борьба с помехами. М., Физматгиз, 1963.

21. **Удалов А. П., Супрун Б. А.** Избыточное кодирование при передаче информации двоичными кодами. М., «Связь», 1964.

22. **Берлекамп Е. Р.** Алгебраическая теория кодирования. М., «Мир», 1972.

23. **Шляпоберский В. И.** Основы техники передачи дискретных сообщений. М., «Связь», 1973.

24. **Мешковский К. А., Кириллов Н. Е.** Кодирование в технике связи. М., «Связь», 1966.

25. **Коды** с обнаружением и исправлением ошибок. Сб. статей. Пер. с англ. Под ред. А. М. Петровского. М., ИЛ, 1956.

26. **Блох Э. Л., Попов О. В., Турин В. Я.** Модели источника ошибок в каналах передачи цифровой информации. М., «Связь», 1971.

27. **Финк Л. М.** Теория передачи дискретных сообщений. М., «Сов. радио», 1970.

28. **Колесник В. Д., Мирончиков Е. Т.** Декодирование циклических кодов. М., «Связь», 1968.

29. **Возенкрафт Д., Рейффен Б.** Последовательное декодирование. М., ИЛ, 1963.

30. **Месси Дж.** Пороговое декодирование. М., «Мир», 1966.

31. **Галлагер Р.** Коды с малой плотностью проверок на четность. М., «Мир», 1966.

32. **Форни Д.** Каскадные коды. М., «Мир», 1970.

33. **Блох Э. Л.** Помехоустойчивость систем связи с переспросом. М., Изд. АН СССР, 1963.

34. **Перри К. Е., Возенкрафт Дж. М.** «SECO» — саморегулирующееся кодирующе — декодирующее устройство с исправлением ошибок. — «Зарубежная радиоэлектроника», 1964, № 9.

35. **Каневский З. М.** Передача сообщений с информационной обратной связью. М., «Связь», 1969.

Глава 7

Расчет основных характеристик цифровых линий связи

7.1. ВВЕДЕНИЕ

При расчете и проектировании цифровых линий связи различного назначения одной из важных задач является установление соотношений между параметрами этих линий и требуемыми показателями качества передачи информации.

Обычно исходными данными для расчета являются: 1) достоверность передачи информации, задаваемая допустимой вероятностью ошибки на один информационный символ (или на кодовое слово); 2) скорость передачи информации; 3) дальность действия линии связи; 4) вид канала связи, характеризуемый условиями распространения сигналов в таком канале, статистикой помех, условиями работы (стационарная или динамическая линия связи).

Кроме перечисленных данных, при расчете и проектировании необходимо учитывать ряд ограничений, налагаемых на некоторые параметры и характеристики линий связи. К ним могут относиться ограничения на вес и габариты передающей и приемной частей системы связи, на диапазон рабочих частот и ширину спектра сигнала и ряд других ограничений, определяемых спецификой конкретных условий работы рассматриваемой линии связи.

Расчет и проектирование линий связи с оптимальными характеристиками представляет очень сложную задачу, которая требует учета ряда показателей качества и минимизации «затрат» для достижения требуемых показателей при наложенных ограничениях. Поскольку в реальных условиях ряд параметров линии связи может непрерыво и случайным образом меняться, задача еще более усложняется. Некоторые вопросы оптимизации линии связи и возникающие при этом трудности были рассмотрены в гл. 1, где было выяснено, что в общем виде задачу оптимизации линий связи решить до сих пор не удается.

Далее излагается методика инженерного расчета линий связи, основанная на энергетическом подходе к определению необходимых параметров линии, и приложение этой методики к некоторым важным случаям.

Рассмотрение проведем для линий связи со свободно распространяющимися сигналами и прямой волной в предположении, что способы передачи и приема, вид канала и статистика помех известны.

7.2. ИНЖЕНЕРНАЯ МЕТОДИКА ЭНЕРГЕТИЧЕСКОГО РАСЧЕТА ЛИНИЙ СВЯЗИ

Установим общие соотношения, связывающие основные параметры линии передачи информации.

В соответствии с известным уравнением дальности связи мощность сигнала на входе приемника определяется выражением

$$P_{\text{с вх}} = P_{\text{изл}} \gamma_E G S_э / 4\pi r^2, \qquad (7.1)$$

где $P_{\text{изл}}$ — мощность сигнала, излучаемого передатчиком (обычно под $P_{\text{изл}}$ понимается средняя мощность излучения); G — коэффициент направленного действия антенны передатчика; $S_э$ — эффективная площадь приемной антенны; r — расстояние между передатчиком и приемником; γ_E — коэффициент, учитывающий потери энергии сигнала в среде за счет поглощения.

Коэффициент потерь γ_E обычно принято выражать в виде

$$\gamma_E = 10^{-0,1\alpha r}, \qquad (7.2)$$

где α — коэффициент затухания, определяющий потери на поглощение в децибеллах на километр.

В ряде случаев удобнее перейти от десятичного основания к натуральному. Тогда

$$\gamma_E = \exp(-0,23\alpha r). \qquad (7.3)$$

Коэффициент затухания α зависит от физико-химических свойств среды распространения и частоты.

Экспериментальные исследования показывают, что для электромагнитных колебаний с длиной волны $\lambda > 10$ см потери поглощения невелики и с ними в первом приближении можно не считаться. Однако для колебаний с длиной волны $\lambda < 5$ см потери поглощения стано-

вятся заметными. Так, например, по данным работы [1] в зависимости от метеорологических условий потери в атмосфере могут составлять

$\alpha \approx 0{,}002$—$0{,}2$ дБ/км для $\lambda \approx 5$ см;

$\alpha \approx 0{,}01$—$1{,}0$ дБ/км для $\lambda \approx 3$ см.

Для акустических сигналов, распространяющихся в воде, затухание, связанное с поглощением энергии сигнала, определяется выражением [2]

$$\alpha \approx 0{,}036 f^{3/2} \text{ [дБ/км]}, \qquad (7.4)$$

где f — частота, выраженная в килогерцах.

Итак, с учетом (7.3) выражение (7.1) принимает вид

$$P_{\text{с вх}} = P_{\text{изл}} \frac{G S_э}{4\pi r^2} \exp(-0{,}23\alpha r). \qquad (7.5)$$

Рассчитывая параметры цифровой линии связи, разработчик системы ориентируется на определенные исходные данные и начальные условия: тип канала (радиоканал, акустический канал и т. п.), диапазон возможных рабочих частот, тип ожидаемых помех и их статистику, вид сигнала и ширину его спектра.

Если основными помехами в линии связи являются внутренние флюктуационные шумы и другие случайные помехи шумового типа (например, космические шумы в радиолиниях, шумы моря в гидроакустических линиях, и т. п.), то, пересчитав все эти помехи ко входу приемника, можно определить результирующую спектральную плотность помех на входе в виде

$$N_{0\Sigma}(f) = \sum_i N_{0i}(f), \qquad (7.6)$$

где N_{0i} — спектральная плотность случайной помехи i-го вида, пересчитанная ко входу приемника.

Мощность всех помех на входе приемника, определяемая в полосе частот $\Delta f_э$, занимаемой спектром сигнала, равна

$$P_{\text{ш вх}} = \int\limits_{f_0 - \Delta f_э/2}^{f_0 + \Delta f_э/2} N_{0\Sigma}(f)\, df, \qquad (7.7)$$

где f_0 — частота несущей.

Выражение (7.7) можно представить в виде

$$P_{\text{швх}} = \overline{N_{0\Sigma}(f)}\, \Delta f_э = N_{0\Sigma} \Delta f_э, \qquad (7.8)$$

где

$$N_{0\Sigma}=\overline{N_{0\Sigma}(f)}=\frac{1}{\Delta f_э}\int\limits_{f_0-\Delta f_э/2}^{f_0+\Delta f_э/2} N_{0\Sigma}(f)\,df \qquad (7.9)$$

можно рассматривать как среднюю спектральную плотность суммарной помехи на входе приемника.

В простейшем случае, когда основной помехой являются только внутренние флюктуационные шумы приемника с равномерной спектральной плотностью N_0, мощность помехи на входе (при согласованном входе) равна

$$P_{ш\,вх}=N_0\Delta f_э=kT_э\Delta f_э, \qquad (7.10)$$

где k — постоянная Больцмана ($k=1{,}38\cdot10^{-23}$ Дж/К), $T_э$ — эквивалентная шумовая температура входа.

С учетом выражений (7.5) и (7.8) отношение средней мощности сигнала к средней мощности шума на входе приемника определяется формулой

$$\left(\frac{P_с}{P_ш}\right)_{вх}=\frac{P_{изл}GS_э}{4\pi r^2 N_{0\Sigma}\Delta f_э}\exp(-0{,}23\alpha r). \qquad (7.11)$$

Это выражение определяет фактическое отношение сигнал/шум на входе приемника при известных параметрах линии связи.

Пусть для того, чтобы обеспечить требуемую вероятность ошибки при передаче одной двоичной единицы информации, необходимо иметь энергетическое отношение сигнал/шум

$$h^2_{0\,тр}=E_0/N_{0\Sigma}=(P_с/P_ш)_{вх}\,\tau_0\Delta f_э. \qquad (7.12)$$

Тогда требуемое отношение сигнал/шум на входе приемника

$$(P_с/P_ш)_{тр}=h^2_{0\,тр}/\tau_0\Delta f_э. \qquad (7.13)$$

При определении требуемого отношения сигнал/шум в соответствии с выражением (7.13) обычно не учитывается ряд причин, снижающих помехоустойчивость приема. К их числу относятся неидеальность синхронизации, наличие интерференционных искажений, влияние трудно учитываемых особенностей распространения и ряд других факторов *.

* Учет указанных причин существенно затрудняет анализ помехоустойчивости и обычно такой анализ проводят при упрощающих допущениях, а затем стремятся обеспечить такие условия работы и так выбрать параметры, чтобы принятые допущения были приемлемы.

Чтобы ослабить влияние указанных неучитываемых причин, часто требуемое отношение сигнал/шум берут с некоторым запасом (системный запас), т. е.

$$(P_с/P_ш)_{тр} = \gamma_{сист} h^2{}_{0\,тр}/\tau_0 \Delta f_э. \qquad (7.14)$$

Коэффициент запаса $\gamma_{сист}$ в зависимости от конкретных условий обычно выбирают в пределах от 2 до 10 (3—10 дБ).

Для того чтобы линия связи обеспечивала передачу информации с помехоустойчивостью не ниже заданной, необходимо выполнить условие

$$(P_с/P_ш)_{вх} \geqslant (P_с/P_ш)_{тр}. \qquad (7.15)$$

Приняв во внимание (7.11), (7.14) и (7.15), имеем

$$\frac{P_{изл} G S_э}{4\pi r^2 N_{0\Sigma}} \exp(-0{,}23\alpha r) \geqslant \gamma_{сист} \frac{h^2{}_{отр}}{\tau_0}. \qquad (7.16)$$

Из теории информации известно, что пропускная способность симметричного дискретного канала * определяется выражением [3]

$$C = \frac{1}{\tau_0}\left[\log_2 n + p_{ош} \log_2 \frac{p_{ош}}{n-1} + (1 - p_{ош}) \log_2 (1 - p_{ош})\right], \qquad (7.17)$$

где n — основание первичного кода (см. гл. 2); $p_{ош}$ — вероятность ошибочного приема информационного символа длительностью τ_0.

При условии, что требуется малая вероятность ошибочного приема ($p_{ош} \ll 1$), выражение (7.17) можно записать в виде

$$C \approx \log_2 n/\tau_0 = \max R_т \log_2 n, \qquad (7.18)$$

где $\max R_т$ — максимальная техническая скорость передачи цифровой информации, выраженная в бодах.

В важном частном случае симметричного бинарного канала ($n=2$) имеем

$$C \approx \max R_т = 1/\tau_0, \qquad p_{ош} \ll 1. \qquad (7.19)$$

В реальных условиях обеспечить передачу со скоростью, определяемой выражениями (7.18) и (7.19), не удается, так как реальная техническая скорость переда-

* Симметричным дискретным называют канал, у которого одинаковы алфавиты на входе и выходе и одинаковы условные вероятности ошибочного приема любого из символов входной цифровой последовательности.

чи снижается из-за того, что часть общего времени передачи приходится тратить на передачу синхронизирующих сигналов, а также на защитные интервалы между кодовыми комбинациями. Поэтому можно записать

$$R=\gamma_R \max R_т=\gamma_R/\tau_0, \qquad (7.20)$$

где γ_R — коэффициент, учитывающий уменьшение скорости передачи. Нетрудно видеть, что $\gamma_R<1$ (см. также § 4.8).

В реальных каналах передачи информации величина γ_R может принимать значения порядка 0,6—0,9 в зависимости от конкретных условий работы, параметров канала и т. п.

С учетом изложенного реальная информационная скорость передачи в симметричном дискретном канале с малой вероятностью ошибки определяется величиной

$$R=\gamma_R \max R_т \log_2 n=\gamma_R \log_2 n/\tau_0. \qquad (7.21)$$

С учетом (7.21) выражение (7.16) принимает вид

$$\frac{P_{изл} G S_э}{4\pi r^2 N_{0\Sigma}} \exp(-0{,}23\alpha r) \geqslant \frac{\gamma_{сист}}{\gamma_R} \frac{h^2_{0\ тр}}{\log_2 n} R. \qquad (7.22)$$

Это выражение является исходным и позволяет решать разнообразные задачи, связанные с расчетом параметров линии связи *.

Приложение изложенной методики расчета линий связи проиллюстрируем некоторыми примерами, которые рассматриваются в следующих параграфах.

7.3. ЭНЕРГЕТИЧЕСКИЙ РАСЧЕТ КОСМИЧЕСКИХ ЛИНИЙ СВЯЗИ

Заранее оговоримся что рассматриваемые ниже расчеты не относятся к какой-либо конкретной линии связи и опираются на некоторые средние исходные данные, ко-

---

* При расчетах различных линий связи часто пользуются децибельной мерой, которая позволяет наглядно и просто оценивать роль тех или иных параметров в энергетике линии. Так, например, выражение (7.5), представленное в децибельной мере $P_{с\ вх}$[дБ·Вт]$=10 \lg P_{с\ вх}$[Вт], принимает вид $P_{с\ вх}=P_{изл}+G+S_э-\alpha r-2r-11$. Здесь все величины выражены в децибелах. Так как децибельная мера является относительной, необходимо следить за тем, чтобы все величины одинаковых или родственных категорий были представлены в одинаковых единицах измерения. Например, $S_э$ и r^2 должны быть представлены в [м^2] или [км^2] и т. п. С применением децибельной меры при расчетах линий связи можно ознакомиться в работах [10—12].

торые характерны для современных космических линий дальней связи и приводятся в работах [4—6].

Основная цель этого параграфа состоит в том, чтобы проиллюстрировать методику расчета и указать примерный порядок параметров космических линий связи. При этом будем иметь в виду, что все космические каналы связи в первом приближении можно рассматривать как каналы гауссовского типа. Это допустимо, поскольку в космических каналах можно не считаться с эффектом многолучевости, а возможные флюктуации сигнала из-за случайных изменений положения антенны космического аппарата на траектории сравнительно невелики и их обычно учитывают при выборе системного коэффициента запаса $\gamma_{сист}$.

Рассмотрение проведем для двух случаев.

7.3.1. Радиолиния «космический аппарат — Земля»

Существующие применяемые линии подобного типа предназначены для передачи информации на Землю с космических аппаратов (КА), используемых для исследования планет Солнечной системы. Для таких линий связи характерны следующие данные: мощность бортового передатчика в режиме непрерывного излучения $P_{изл} \approx 25—50$ Вт; рабочая частота $f_0 \approx 2300$ МГц*; диаметр бортовой передающей антенны $D_б \approx 2{,}5—5$ м; диаметр антенны наземной станции $D_з \approx 25—70$ м; эффективная шумовая температура наземного приемника $T_э \approx$ $\approx (25—50)$ К [4—6, 9].

Допустим, что по такой радиолинии необходимо передать телеметрическую информацию с заданной вероятностью ошибки на один информационный символ $p_{ош}$. Будем полагать, что передача осуществляется двоичными сигналами методом ОФМн. При высоких требованиях к достоверности приема информации вероятность ошибки приема таких сигналов можно определить выражением (4.58) или (4.59). Как показано в гл. 4, эти выражения приблизительно равноценны. Поэтому будем пользовать-

---

* Рабочие частоты для систем дальней космической связи установлены международными соглашениями и выбираются в диапазоне (2290—2300) МГц для линий «космический аппарат — Земля» и в диапазоне (2110—2120) МГц для линий «Земля — космический аппарат» [4, 5, 10].

ся выражением (4.59) как более простым. В соответствии с ним можно записать

$$p_{\text{ош}}=0{,}5\exp(-h^2{}_0),$$

отсюда следует

$$h^2{}_{0\,\text{тр}}\geqslant h^2{}_0=\ln(1/2p_{\text{ош}}). \tag{7.23}$$

Чтобы выполнить расчеты в соответствии с основным выражением (7.22), кроме величины $h_{\text{отр}}$ нужно найти эффективную площадь приемной антенны и коэффициент направленного действия передающей антенны. Из теории антенн известно, что коэффициент направленного действия антенны определяется выражением

$$G_A=\frac{4\pi S_A}{\lambda^2}\eta_{\text{А}}=\frac{4\pi S_{\text{э}}}{\lambda^2}, \tag{7.24}$$

где $\eta_{\text{А}}$ — коэффициент, учитывающий эффективность использования общей площади раскрыва антенны $S_{\text{А}}$ (апертуры антенны); $S_{\text{э}}$ — эффективная площадь раскрыва антенны.

Величина коэффициента $\eta_{\text{А}}$ зависит от типа и конструкции антенны. Для антенн параболического типа $\eta_{\text{А}}\approx 0{,}5$—$0{,}7$ [7]. При расчетах величину $\eta_{\text{А}}$ обычно берут равной 0,55. Выразив площадь антенны $S_{\text{А}}$ через диаметр D, нетрудно получить

$$G=\eta_{\text{А}}(\pi D/\lambda)^2, \tag{7.25}$$

$$S_{\text{э}}=\eta_{\text{А}}\pi D^2/4. \tag{7.26}$$

Ширина диаграммы направленности антенны по уровню половинной мощности определяется приближенной формулой

$$\theta\approx 70\lambda/D. \tag{7.27}$$

Принимая во внимание (7.23), (7.25) и (7.26) и учитывая, что поглощение энергии сигнала в атмосфере на рабочей частоте незначительно и им можно пренебречь, после несложных преобразований выражение (7.22) для двоичных сигналов можно представить в виде

$$\left(\frac{\pi D_{\text{б}} D_{\text{з}}\eta_A}{4\lambda r}\right)^2\frac{P_{\text{изл}}}{N_0}\geqslant\frac{\gamma_{\text{сист}}}{\gamma_R}\ln\left(\frac{1}{2p_{\text{ош}}}\right)R, \tag{7.28}$$

где $\lambda=c/f_0$ — длина волны; c — скорость распространения электромагнитных колебаний ($c=3\cdot 10^8$ м/с); N_0 — спектральная плотность флюктуационных шумов на входе приемника.

Соотношение (7.28) позволяет определить любой из параметров линии связи при условии, что все остальные параметры известны.

Если учесть, что для рассматриваемой линии связи

$$\lambda = c/f_0 = 3\cdot 10^8/2{,}3\cdot 10^9 \approx 0{,}13 \text{ м и } N_0 \approx$$

$$\approx 1{,}38\cdot 10^{-23}\cdot 50 \approx 0{,}7\cdot 10^{-21} \text{ [Вт/Гц], и принять } \gamma_{сист} = 4,$$

$$\gamma_R = 0{,}75,\ \eta_A \approx 0{,}55,\ D_б = 5 \text{ м},\ D_з = 60 \text{ м},\ P_{изл} = 30 \text{ Вт},$$

то для заданных конкретных значений параметров из (7.28) следует соотношение

$$r^2 R \ln (^1/_2 p_{ош}) \leqslant 0{,}8\cdot 10^{22}. \qquad (7.29)$$

Здесь расстояние r берется в километрах. При заданном значении вероятности ошибки $p_{ош}$ выражение (7.29) определяет связь между достижимой дальностью связи и возможной при этом скоростью передачи.

В линиях дальнеи космической связи отношение сигнал/шум на входе приемника меняется в соответствии с изменением расстояния между Землей и космическим аппаратом. Для поддержания постоянства вероятности ошибки $p_{ош}$ необходимо изменять скорость передачи R в соответствии с изменением расстояния r, т. е. иметь адаптивную линию связи. Практически подобная адаптация достигается с помощью специального программного устройства дискретно управляющего скоростью передачи. При этом производятся соответствующие изменения и в приемном устройстве (дискретно изменяется полоса пропускания, постоянные времени интеграторов и т. п.). Обычно такое изменение делают кратным двум.

Расчеты показывают, что современные применяемые линии дальней космической связи могут обеспечивать довольно высокие скорости передачи на расстояниях до нескольких сотен миллионов километров. Необходимо отметить, что практическая реализация систем с такими показателями является весьма сложной технической задачей. Она связана с созданием малошумящих приемных устройств, применением эффективных методов передачи и приема сигналов, созданием высокоэффективных приемных антенн и увеличением излучаемой мощности. Обсуждение подобных вопросов можно найти в работах [8, 9, 15].

7.3.2. Радиолиния «Земля — космический аппарат»

Линия связи подобного типа предназначена для передачи различных команд с Земли на борт космического аппарата с целью включения или выключения программных бортовых устройств, коррекции траектории полета аппарата и т. п. Достоверность приема таких команд должна быть весьма высокой, а допустимая вероятность ошибки принятой команды составляет 10^{-5}—10^{-6} и менее.

Для современных применяемых линий связи «Земля — космический аппарат» характерны следующие параметры мощность наземного передатчика в режиме непрерывного излучения $P_{изл} \approx 10^3$—10^5 Вт; рабочая частота $f_0 \approx 2100$ МГц; диаметр передающей наземной антенны $D_з \approx 25$ м; диаметр бортовой приемнои антенны $D_б = 2{,}5$—5 м; эффективная шумовая температура бортового приемника $T_э \approx 500$—1000 К [4—6, 9].

Если принять $P_{изл} = 10^3$ Вт; $D_б = 5$ м; $T_э = 1000$ К; $D_з = 25$ м; $\gamma_{сист} = 4$; $\gamma_R = 0{,}75$; $\eta_A = 0{,}55$, то в соответствии с выражением (7.28) получим

$$r^2 R \ln (1/2p_{ош}) \leqslant 0{,}2 \cdot 10^{22}. \quad (7.30)$$

Нетрудно убедиться в том, что радиолиния с указанными параметрами может обеспечить передачу команд на весьма значительные расстояния. Так, например, при допустимой вероятности ошибки $p_{ош} = 10^{-6}$ и скорости передачи информации о команде $R = 8$ бит/с, дальность действия радиолинии составит около 4,5 млрд. км.

7.4. ЭНЕРГЕТИЧЕСКИЙ РАСЧЕТ ЦИФРОВОЙ ГИДРОАКУСТИЧЕСКОЙ ЛИНИИ СВЯЗИ

Возрастающий интерес во многих странах к исследованию морей и океанов привел к быстрому развитию гидроакустики как в научном, так и в прикладном направлениях [12—14]. В частности, большое внимание в последние годы уделяется разработке методов передачи и приема информации под водой и созданию средств гидроакустическои связи для океанологических исследований, разведки и добычи полезных ископаемых на дне морей и океанов и т. д. [13, 14].

Распространение акустических сигналов в море сопровождается рядом сложных явлений, обусловленных отражениями от поверхности моря и дна, рассеянием на неоднородностях и поглощением энергии сигналов в морской среде. Эти явления приводят к тому, что параметры посылок гидроакустических сигналов, приходящих на вход приемника, приобретают случайные изменения, форма огибающей каждой посылки искажается, между посылками появляется интерференция, причиной которой являются как указанные искажения, так и многолучевой характер распространения.

На прием сигналов в гидроакустических каналах связи могут оказывать существенное влияние различные помехи: собственные шумы моря, шумы обитателей моря, шумы судовых двигателеи и т. п. [2, 13]. Учесть перечисленные выше факторы при расчете гидроакустической линии связи в настоящее время не представляется возможным, так как многие из них до сих пор еще недостаточно изучены.

Дальнейшее рассмотрение проведем при следующих условиях и допущениях:

1. Для передачи цифровой информации по гидроканалу применяются простые двоичные сигналы, инвариантные к частотным свойствам морской среды. Это означает, что длительность посылки сигнала выбирается так, чтобы неравномерность коэффициента затухания в пределах диапазона частот, занимаемого спектром посылки Δf_c, была незначительной, и затухание можно было бы полагать одинаковым для всех частот спектра сигнала. Сделанное допущение эквивалентно требованию, чтобы рассматриваемый гидроакустический канал был узкополосным, т. е. выполнялось условие $(\Delta f_c/f_0) \ll 1$, где f_0 — несущая частота посылки сигнала.

Допущение об узкополосности канала позволяет считать, что среда не искажает форму огибающей посылки сигнала, пришедшего на вход приемника, а уменьшение энергии посылки из-за поглощения в морской среде определяется значением коэффициента затухания на несущей частоте посылки.

2. Считается, что влияние многолучевого эффекта при распространении акустических сигналов в морской среде незначительно и его можно не учитывать. Это допустимо, если дальность связи сравнительно невелика (не более 2—4 км) и приняты меры к уменьшению влияния отра-

жений сигнала от поверхности моря и дна. Указанные условия работы характерны для современных и перспективных информационных гидроакустических систем [13].

3. Из различного вида возможных помех учитываются только принципиально неустранимые собственные шумы моря. Эти шумы обусловлены тепловыми явлениями, происходящими в воде, и разнообразными факторами механического происхождения (волнением поверхности моря, брызгами, лопающимися пузырьками воздуха в приповерхностном слое воды и т. п.). Уровень шумов моря существенно зависит от состояния поверхности и метеорологических условий. Как правило, уровень шумов моря значительно превышает уровень внутренних шумов приемника, поэтому влияние последних можно не учитывать. Учет других видов помех затруднителен, поскольку эти помехи изучены еще недостаточно. В ряде случаев их действие можно существенно ослабить правильным выбором параметров системы связи (рабочей частоты сигнала и т. п.) и условий работы системы.

Второе и третье допущения позволяют считать, что в первом приближении рассматриваемый узкополосный гидроакустический канал связи является гауссовским.

Суть энергетического расчета цифровой гидроакустической линии связи состоит в том, чтобы найти отношение сигнал/помеха на входе приемника, при котором обеспечивается требуемое качество передачи информации.

В соответствии с уравнением дальности связи (7.5) средняя мощность акустического сигнала на входе приемника, приходящаяся на единицу эффективной площади приемной антенны (средняя интенсивность сигнала), равна

$$I_{\text{свх}} = \frac{P_{\text{свх}}}{S_{\text{э}}} = P_{\text{изл}} \frac{G_1}{4\pi r^2} \exp(-0{,}23\alpha r) \left[\frac{\text{Вт}}{\text{м}^2}\right], \quad (7.31)$$

где G_1 — коэффициент направленного действия излучателя передатчика (коэффициент концентрации); $S_{\text{э}}$ — эффективная площадь приемной антенны (гидрофона). Остальные обозначения те же, что и в выражении (7.5).

Коэффициент затухания акустических колебаний в морской среде α определяется выражением (7.4). При теоретических исследованиях и расчетах часто удобно аппроксимировать коэффициент затухания линейной или квадратичной функцией частоты. Запишем аппроксимацию коэффициента затухания в виде

$$\alpha = \beta(n) f^n \ [\text{дБ/км}], \quad (7.32)$$

где n — показатель, величина которого в зависимости от применяемой аппроксимации, может быть выбрана в пределах $1 \leqslant n \leqslant 2$; $\beta(n)$ — постоянный коэффициент, значение которого определяется величиной выбранного показателя n. Естественно, что при выбранной величине n нужно подобрать такое значение коэффициента $\beta(n)$, чтобы аппроксимация вида (7.32) соответствовала экспериментальной зависимости $\alpha(f)$ в рассматриваемом диапазоне частот $(f_1—f_2)$.

Если для одного из видов аппроксимации коэффициент $\beta(n)$ известен, несложно определить его значение для другого вида аппроксимации. Можно показать, что минимальная средквадратичная ошибка при переходе от аппроксимации с показателем n_1 к аппроксимации с показателем n_2 достигается при выборе коэффициента β по формуле

$$\beta(n_2) = \beta(n_1)\, f_2^{n_1 - n_2}, \tag{7.33}$$

где f_2 — верхняя граничная частота используемого диапазона.

В частности, если исходить из аппроксимации вида (7.4), то для линейной и квадратичной аппроксимации имеем соответственно

$$\beta(1) = \beta(3/2)\sqrt{f_2} \approx 0{,}036\sqrt{f_2}, \tag{7.34}$$

$$\beta(2) = \beta(3/2)/\sqrt{f_2} \approx 0{,}036/\sqrt{f_2}. \tag{7.35}$$

Приводимые в ряде работ по гидроакустике [2, 13, 14] экспериментальные зависимости спектральной плотности интенсивности собственных шумов моря от частоты при разных значениях волнения моря (кривые Кнудсена) можно аппроксимировать выражением

$$N_{0м}(f) = A(b)/f^2 + Bf^2, \tag{7.36}$$

где $A(b)$ — коэффициент, величина которого зависит от состояния поверхности моря b, определяемого в баллах; B — коэффициент, величина которого зависит от температуры морской среды.

Расчеты показывают, что если частоту выражать в килогерцах, то эти коэффициенты имеют следующие значения:

$$A(b) \approx 2{,}5 \cdot 10^{-8} \exp[-5{,}3 \exp(-0{,}6b)], \tag{7.37}$$

$$B \approx 2{,}2 \cdot 10^{-17} \text{ для } t = 15°\text{C}. \tag{7.38}$$

Спектральная плотность интенсивности шумов моря при этом имеет размерность [Вт/м²·кГц]. Заметим, что указанные выше значения коэффициентов ориентировочны и должны уточняться экспериментально для каждого конкретного района.

Шумы моря в первом приближении можно считать изотропными [2]. Если приемный гидрофон имеет коэффициент направленности G_2, то эти шумы, приведенные к входу приемника, ослабляются в G_2 раз. Учитывая сказанное, среднюю интенсивность шумов моря на входе приемника в полосе частот Δf_c, занимаемой спектром посылки сигнала, можно определить выражением

$$J_{\text{швх}} = \int\limits_{f_0-\Delta f_c/2}^{f_0+\Delta f_c/2} \frac{N_{0M}(f)}{G_2} df, \qquad (7.39)$$

где

$$f_0 = 0{,}5(f_1 + f_2), \quad \Delta f_c = f_2 - f_1. \qquad (7.40)$$

Учитывая допущение об узкополосности рассматривае мого канала ($\Delta f_c/f_0 \ll 1$) и приняв во внимание выражение (7.36), после несложных вычислений имеем

$$J_{\text{швх}} \approx \frac{\Delta f_c}{G_2} [A(b)/f_0^2 + Bf_0^2]. \qquad (7.41)$$

При получении выражения (7.41) учтено условие $\Delta f_c/f_0 \ll$ $\ll 1$ и сделаны соответствующие пренебрежения малыми второго порядка. Выражение (7.41) можно представить в виде

$$J_{\text{швх}} = \overline{N_{0M}(f)}\, \Delta f_c, \qquad (7.42)$$

где

$$\overline{N_{0M}(f)} = G_2^{-1} [A(b)/f_0^2 + Bf_0^2]. \qquad (7.43)$$

Для рассматриваемого диапазона частот величина $\overline{N_{0M}(f)}$ определяет среднюю спектральную плотность интенсивности шумов моря, пересчитанную к входу приемника.

Приняв во внимание соотношения (7.31), (7.32) и (7.41), можно записать следующее выражение для фактического отношения мощности сигнала к мощности шума на входе приемника гидроакустической линии связи:

$$\left(\frac{P_c}{P_{\text{ш}}}\right)_{\text{вх}} = \left(\frac{J_c}{J_{\text{ш}}}\right)_{\text{вх}} = \frac{P_{\text{изл}} G_1 G_2 \exp[-0{,}23\beta(n) f_0^n r]}{4\pi r^2 \Delta f_c [A(b)/f_0^2 + Bf_0^2]}. \qquad (7.44)$$

Располагая выражением (7.44) и поступая в соответствии с методикой, изложенной в § 7.2 (см. (7.16)), можно записать условие

$$\frac{P_{\text{изл}} G_1 G_2 \exp[-0{,}23\beta(n) f_0^n r]}{4\pi r^2 [A(b)/f_0^2 + B f_0^2]} \geqslant \gamma_{\text{сист}} \frac{h_{0\,\text{тр}}^2}{\tau_0}, \qquad (7.45)$$

где $h^2_{0\,\text{тр}}$ — энергетическое отношение сигнал/шум, при котором обеспечивается требуемая вероятность ошибки при передаче одной двоичной единицы информации; $\gamma_{\text{сист}}$ — системный коэффициент запаса, учитывающий факторы, количественная оценка влияния которых затруднительна (межсимвольные помехи, неидеальность синхронизации, случайные изменения условий распространения и т. д.).

При пользовании выражением (7.45) необходимо иметь в виду, что длительность посылки τ_0 акустического сигнала не может выбираться произвольно и должна удовлетворять принятому допущению об узкополосности рассматриваемого гидроакустического канала связи. Найдем условия справедливости такого допущения. Экспоненциальный множитель в выражении (7.31) учитывает частотные свойства морской среды и, по сути дела, представляет собой коэффициент передачи среды по мощности. Будем понимать под узкополосным гидроакустическим каналом связи канал с полосой частот Δf, в пределах которой неравномерность коэффициента передачи среды незначительна. Количественно эту неравномерность можно оценить величиной

$$\eta_{\Delta f} = \frac{K_c(f_2)}{K_c(f_1)} = \frac{\exp[-0{,}23\beta(n) f_2^n r]}{\exp[-0{,}23\beta(n) f_1^n r]}, \qquad (7.46)$$

где f_2 и f_1 — верхняя и нижняя границы диапазона частот, занимаемого гидроканалом; $K_c(f)$ — коэффициент передачи среды по мощности на частоте f.

Нетрудно видеть, что неравномерность коэффициента передачи заключена в пределах $0 \leqslant \eta_{\Delta f} \leqslant 1$. Чем ближе величина $\eta_{\Delta f}$ к единице, тем меньше неравномерность. Для узкополосного канала неравномерность $\eta_{\Delta f}$ не должна быть менее 0,5, т. е. не должна превышать 3 дБ.

Если принять квадратичную аппроксимацию коэффициента затухания, несложно показать, что условие узко-

полосности канала определяется выражением

$$\Delta f \leqslant 2{,}2 \frac{\ln(1/\eta_{\Delta f})}{\beta(2) f_0 r}, \quad 0{,}5 \leqslant \eta_{\Delta f} \leqslant 1, \qquad (7.47)$$

где $\Delta f \leqslant f_2 - f_1$, $f_0 = 0{,}5(f_2 + f_1)$.

При выборе ширины спектра посылок двоичного сигнала в соответствии с условием (7.47), посылки не будут существенно искажаться из-за частотных свойств среды. Из выражения (7.47) вытекает условие выбора длительности посылки сигнала, инвариантного к частотным свойствам среды:

$$\tau_0 \geqslant \frac{k_ф}{\Delta f} \approx 0{,}46 \frac{k_ф \beta(2) f_0 r}{\ln(1/\eta_{\Delta f})}, \quad 0{,}5 \leqslant \eta_{\Delta f} \leqslant 1, \qquad (7.48)$$

где $k_ф$ — коэффициент, зависящий от вида манипуляции сигнала и допустимой степени искажения посылки на выходе канала. При проведении анализа обычно принимают для АМн и ФМн сигналов $k_ф = 1$, а для ЧМн $k_ф = 2$.

Сигнал, длительность посылки которого выбрана в соответствии с условием (7.48), называют инвариантным к частотным свойствам среды. С учетом условия (7.48) для квадратичной аппроксимации выражение (7.45) принимает вид

$$\frac{P_{изл} G_1 G_2 \exp[-0{,}23\beta(2) f_0^2 r]\, k_ф \beta(2) f_0 r}{8{,}7\pi r^2 [A(b)/f_0^2 + B f_0^2] \ln(1/\eta_{\Delta f})} \geqslant \gamma_{сист} h^2_{0\,тр}. \qquad (7.49)$$

Величина $h^2_{0тр}$ определяет требуемое энергетическое отношение сигнал/шум с учетом заданного вида манипуляции сигнала и выбранного способа приема. Ее значение можно найти из выражения для вероятности ошибки, наиденного с учетом применяемых способов передачи и приема и типа канала. В частности, для простых двоичных сигналов и некогерентного приема в гауссовском канале это выражение имеет вид (см. гл. 4)

$$p_{ош} = 0{,}5 \exp(-\gamma^2_с h^2_{0\,тр}/2),$$

отсюда следует

$$h^2_{0\,тр} = (2/\gamma^2_с) \ln(1/2 p_{ош}). \qquad (7.50)$$

К гауссовским каналам можно отнести гидроакустические каналы связи со сравнительно небольшой дальностью действия ($r \approx 1$—2 км), в которых приняты меры для существенного ослабления влияния отражений сиг-

нала от поверхности и дна. Каналы со значительно большей дальностью действия, в которых заметную роль начинают играть неоднородности среды и отражения от поверхности и дна, нельзя считать гауссовскими. К сожалению, определение требуемой величины $h^2_{0тр}$ в таких каналах может натолкнуться на значительные трудности, связанные с нахождением выражения для вероятности ошибки. Для получения ориентировочных верхних оценок подобные каналы можно считать релеевскими.

Располагая приведенными выше зависимостями, можно проводить расчеты различных параметров узкополосных цифровых гидроакустических каналов связи.

Остановимся кратко на особенностях расчета широкополосных гидроакустических каналов связи, в которых неравномерность коэффициента передачи среды в пределах занимаемой полосы частот может быть весьма значительной ($\eta_{\Delta f} \ll 1$). При передаче цифровой информации по таким каналам общий занимаемый диапазон частот обычно разбивается на ряд узкополосных поддиапазонов, в каждом из которых информация передается на своей несущей частоте [13, 14]. Иными словами, в широкополосных гидроакустических каналах применяют частотное уплотнение, благодаря которому такие каналы заменяются совокупностью узкополосных каналов *. Этот принцип позволяет более просто решить ряд технических проблем, связанных с созданием сравнительно высокоинформативных гидроакустических систем связи и передачи данных. Если параметры частотного уплотнения (частоты поднесущего, число каналов и т. п.) заданы или выбраны, то дальнейший расчет сводится к определению параметров отдельных узкополосных каналов в соответствии с проведенным выше рассмотрением.

КОНТРОЛЬНЫЕ ВОПРОСЫ И ЗАДАЧИ

7.1. Какие конкретные задачи приходится решать при расчете цифровых линий связи? В чем заключаются основные трудности решения этих задач?

7.2. Назвать основные пути уменьшения системного запаса в линиях связи. Сформулировать критерий целесообразности уменьшения системного запаса.

---

* Подобный принцип уплотнения широко применяется в системах передачи данных, использующих телефонные каналы. Он обеспечивает хорошую эффективность использования полосы (см. гл. 6).

7.3. В космической линии связи «КА — Земля» необходимо увеличить скорость передачи вдвое. Какие мероприятия нужно для этого выполнить? Что изменится, если указанную задачу нужно решить для линии «Земля — КА»?

7.4. Найти скорость передачи по линии связи «КА — Земля» с параметрами: $P_{изл}=3$ Вт; $D_б=1,2$ м; $f_0=960$ мГц; $D_з=16$ м; $T_э=250$ К; $r=2\cdot10^8$ км. Принять $\gamma_{сист}=4$; $\gamma_к=0,7$. Способ передачи — однократная ОФМн, способ приема — когерентный. Допустимая вероятность ошибки $p_{ош}=10^{-4}$.

7.5. Найти возможную скорость передачи в линии «КА — Земля» с параметрами, приведенными в § 7.3 для случаев:
а) связь «район Марса — Земля» (расстояние «Земля — Марс» в зависимости от расположения планет может при-приблизительно меняться от 60 до 380 млн. км);
б) связь «район Юпитера — Земля» (возможные расстояния от 600 до 930 млн. км).

СПИСОК ЛИТЕРАТУРЫ

1. **Калашников Н. И.** Основы электромагнитной совместимости систем связи через ИСЗ с другими службами. М., «Связь», 1970.
2. **Сташкевич А. П.** Акустика моря. Л., «Судостроение», 1966.
3. **Финк Л. М.** Теория передачи дискретных сообщений. М., «Сов. радио», 1970.
4. **Бандик У. Т.** и др. Система связи космического аппарата «Лунар Орбитер». — «Зарубежная радиоэлектроника», 1968, № 5.
5. **Сиверс А. П.** Основы космической радиоэлектроники. М., «Сов. радио», 1969.
6. **Уорд Р. Б.** Применение корреляционного метода радиолокации к задачам связи и слежения за объектами в дальнем космосе. — «Зарубежная радиоэлектроника», 1964, № 12.
7. **Покрас А. М.** Антенные устройства зарубежных линий связи через ИСЗ. М., «Связь», 1965.
8. Цифровые методы в космической связи. Пер. с англ. М., «Связь», 1969.
9. **Гольдштейн Б. С.** Связь с космическим кораблем в районе планеты Марс: требования и ограничения. — «Зарубежная радиоэлектроника», 1968, № 12.
10. **Фортушенко А. Д.** и др. Основы технического проектирования систем связи через ИСЗ. М., «Связь», 1970.
11. **Крэсснер Г. Н., Михаелс Дж.** Введение в системы космической связи. Пер. с англ. М., «Связь», 1967.
12. Подводная акустика. Пер. с англ. М., «Мир», 1970.
13. **Тарасюк Ю. Ф., Серавин Г. Н.** Гидроакустическая телеметрия. Л., «Судостроение», 1973.
14. Достижения в области телеметрии. Пер. с англ. М., «Мир», 1970.
15. **Тепляков И. М., Калашников И. Д., Рощин Б. В.** Радиолинии космических систем передачи информации. М., «Сов. радио», 1975.

Приложение 1

А. Значения функции

$$V(x)=\frac{1}{\sqrt{2\pi}}\int_{x}^{\infty}\exp\left(-\frac{y^2}{2}\right)dy;\ V(x)=0{,}5-\Phi(x).$$

x	0,5	1,0	1,5	1,75	2,0	2,25	2,5
$V(x)$	$3{,}085\cdot10^{-1}$	$1{,}587\cdot10^{-1}$	$6{,}681\cdot10^{-2}$	$4{,}006\cdot10^{-2}$	$2{,}275\cdot10^{-2}$	$1{,}222\cdot10^{-2}$	$6{,}21\cdot10^{-3}$
x	2,75	3,0	3,25	3,5	3,75	4,0	4,25
$V(x)$	$2{,}98\cdot10^{-3}$	$1{,}35\cdot10^{-3}$	$5{,}77\cdot10^{-4}$	$2{,}326\cdot10^{-4}$	$0{,}884\cdot10^{-4}$	$3{,}167\cdot10^{-5}$	$1{,}07\cdot10^{-5}$
x	4,5	4,75	5,0	5,25	5,50	5,75	6,0
$V(x)$	$3{,}4\cdot10^{-6}$	$1{,}017\cdot10^{-6}$	$2{,}866\cdot10^{-7}$	$7{,}605\cdot10^{-8}$	$1{,}9\cdot10^{-8}$	$4{,}462\cdot10^{-9}$	$9{,}866\cdot10^{-10}$

Б. Приближенные значения аргумента x, при которых функция $V(x)=10^{-k}$, где $k=1{,}2\ldots$

$V(x)$	10^{-1}	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
x	1,3	2,32	3,1	3,72	4,27	4,76
$V(x)$	10^{-7}	10^{-8}	10^{-9}	10^{-10}	10^{-11}	10^{-12}—10^{-13}
x	5,2	5,2	6,0	6,37	6,71	7,04—7,35

Приложение 2

Значения функций exp (— x) и ln x.

x	0,5	1,0	1,5	2,0	2,5	3,0	3,5
exp(—x)	$6,065 \cdot 10^{-1}$	$3,68 \cdot 10^{-1}$	$2,23 \cdot 10^{-1}$	$1,353 \cdot 10^{-1}$	$0,821 \cdot 10^{-1}$	$5 \cdot 10^{-2}$	$3 \cdot 10^{-2}$
ln x	—0,69	0	0,405	0,693	0,916	1,1	1,253
x	4,0	4,5	5,0	5,5	6,0	6,5	7,0
exp (—x)	$1,83 \cdot 10^{-2}$	$1,11 \cdot 10^{-2}$	$6,74 \cdot 10^{-3}$	$4,1 \cdot 10^{-3}$	$2,48 \cdot 10^{-3}$	$1.5 \cdot 10^{-3}$	$9,12 \cdot 10^{-4}$
ln x	1,386	1,504	1,609	1,705	1,792	1,872	1,946
x	7,5	8,0	8,5	9,0	9,5	10	10,5
exp (—x)	$5,53 \cdot 10^{-4}$	$3,35 \cdot 10^{-4}$	$2,03 \cdot 10^{-4}$	$1,23 \cdot 10^{-4}$	$7,5 \cdot 10^{-5}$	$4,5 \cdot 10^{-5}$	$2,8 \cdot 10^{-5}$
ln x	2,015	2,079	2,14	2,197	2,25	2,303	2,351
x	12	14	16	18	20	22	24
exp (—x)	$6,3 \cdot 10^{-6}$	$8,7 \cdot 10^{-7}$	$1,18 \cdot 10^{-7}$	$1,6 \cdot 10^{-8}$	$2,14 \cdot 10^{-9}$	$2,8 \cdot 10^{-10}$	$9,1 \cdot 10^{-11}$
ln x	2,485	2,64	2,773	2,89	2,996	3,09	3,18

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

ОГЛАВЛЕНИЕ

ГОТОВЯТСЯ К ВЫХОДУ В ИЗДАТЕЛЬСТВЕ «СОВЕТСКОЕ РАДИО»

Воробьев Е. И., Китов А. И. **Автоматизация обработки информации и управления в здравоохранении.**

Обобщается отечественный и зарубежный опыт разработки и применения автоматизированных систем сбора и обработки информации в медицине и здравоохранении.

Приводятся алгоритмы решения типовых задач (обработка данных о деятельности больничных стационаров, учет и планирование кадров, анализ эпидемиологической обстановки, материально-техническое медицинское снабжение и др.).

А. И. Китов известен первыми в СССР книгами по ЭВМ, а также книгами по программированию информационных, экономических и управленческих задач, Е. И. Воробьев — автор нескольких книг по медицине.

Книга предназначена для специалистов, занимающихся разработкой и применением АСУ и вычислительной техники в медицине и здравоохранении, а также студентов и аспирантов.

Мудров В. И., Кушко В. Л. Методы обработки измерений.

В предлагаемой книге рассматриваются методы определения неизвестных параметров по данным измерений. Описываются различные модели и законы распределения ошибок, возникающих в рамках этих моделей. Особое внимание уделяется причинам возникновения и способам учета так называемой ошибки объекта, обусловленной неточностью математического описания изучаемого явления. Описываются свойства оценок, полученных в результате минимизации ряда функционалов как квадратического, так и неквадратического вида. Много места отводится оценкам, основанным на минимизации суммы модулей уклонений. Приводятся удобные алгоритмы получения оценок и их ковариационных матриц. Использование этих методов иллюстрируется на конкретных примерах.

Книга рассчитана на инженеров широкого профиля.

Уилсон А., Уилсон М. Управление и творчество при проектировании систем. Пер. с англ.

Книга посвящена системному анализу процессов творчества. В современных условиях творчество представляет собой органическое сочетание двух относительно независимых видов деятельности — собственно творческой и управленческой. Авторы доказывают принципиальную возможность управления процессами творчества. Последовательно разбирая эти процессы и представляя их в виде совокупности достаточно стандартизованных формальных приемов, авторы не только определяют задачи как творческого, так и управленческого персонала на каждой стадии, но и предлагают возможные способы их решения.

Книга предназначена для работников любого уровня управления и должна помочь им осознать свое место и роль в процессе творчества, чтобы влиять на него.

Перечисленные книги можно заказать и купить в местных магазинах, распространяющих научно-техническую литературу.

ПЕТР ИЛЬИЧ ПЕНИН

Системы передачи цифровой информации

Редактор **Н. Г. Давыдова**
Художественный редактор **З. Е. Вендрова**
Обложка художника **Б. К. Шаповалова**
Технический редактор **Г. А. Мешкова**
Корректор **Л. А. Максимова**

Сдано в набор 6/X-1975г. Подписано в печать 23/II 1976г. Т-02860
Формат 84×108/32 Бумага типографская № 2
Объем 19,32 усл.-п. л. 19,736 уч.-изд. л.
Тираж 18000 экз. Зак. 376 **Цена 83 к.**

Издательство «Советское радио», Москва, Главпочтамт, а/я 693

Московская типография № 10 Союзполиграфпрома
при Государственном комитете Совета Министров СССР
по делам издательств, полиграфии и книжной торговли.
Москва, М-114, Шлюзовая наб., 10.

Пенин П. И.

П 25 Система передачи цифровой информации. (Учебное пособие для вузов). М., «Сов. радио», 1976.

368 с. с ил.

В книге излагаются принципы, методы и способы создания систем передачи цифровой информации. Большое внимание уделяется постановке задач и получению соотношений для расчетов, а также особенностям передачи цифровой информации по каналам как с постоянными, так и со случайными параметрами.

Книга является учебным пособием для студентов радиотехнических специальностей вузов, а также может быть полезна специалистам в области теории и техники передачи цифровой информации.

П $\frac{30501\text{-}029}{046(01)\text{-}76}$ 72-76 **6Ф7**

Цена 83 к.

